

INFORMATIONSVERVERHALTEN

UND

EFFIZIENZ

Habilitationsschrift

eingereicht bei der

Wirtschafts- und Sozialwissenschaftlichen

Fakultät der

Christian-Albrechts-Universität zu Kiel

zur Erlangung der

Venia Legendi für Betriebswirtschaftslehre

von

Hans Georg Gemünden

Dipl. Kfm. Dr. rer. oec.

aus Ingelheim

Gettorf 1986

INHALTSÜBERSICHT

1. TEIL: PROBLEMSTELLUNG

2. TEIL: DETERMINANTEN DER EFFIZIENZ INDIVIDUELLER ENTSCHEIDUNGEN – EINE SYSTEMATISCHE BESTANDSAUFNAHME EMPIRISCHER FORSCHUNGSERGEBNISSE

2.1. ZIELSETZUNG UND METHODIK DER BESTANDSAUFNAHME

2.2. INFORMATIONEN-SYSTEM UND EFFIZIENZ

2.3. PERSÖNLICHKEIT UND EFFIZIENZ BETRIEBLICHER ENTSCHEIDUNGS-PROZESSE

2.4. INFORMATIONSVRHALTEN UND EFFIZIENZ

3. TEIL: MESSWERTE DER EFFIZIENZ VON BILANZANALYSEN – THEORETISCHES KONZEPT UND EMPIRISCHE VALIDIERUNG –

3.1. THEORETISCHE GRUNDLEGUNG

3.2. OPERATIONALISIERUNG DER TREFFSICHERHEIT DER LAGEBEURTEILUNG

3.3. OPERATIONALISIERUNG DER ANDEREN EFFIZIENZDIMENSIONEN UND DER BEURTEILUNGSTENDENZ

3.4. EMPIRISCHE ANALYSE UND VALIDIERUNG DES MESSKONZEPTES

4. TEIL: EMPIRISCHE UNTERSUCHUNG AM BEISPIEL VON BILANZANALYSEN

4.1. THEORETISCHER BEZUGSRAHMEN DER EMPIRISCHEN UNTERSUCHUNG

4.2. METHODIK UND DURCHFÜHRUNG DER EMPIRISCHEN UNTERSUCHUNG

4.3. DIE ERGEBNISSE DER EMPIRISCHEN UNTERSUCHUNG

1. TEIL: PROBLEMSTELLUNG

2. TEIL: DETERMINANTEN DER EFFIZIENZ INDIVIDUELLER ENTSCHEIDUNGEN
EINE SYSTEMATISCHE BESTANDSAUFNAHME EMPIRISCHER FORSCHUNGSERGEBNISSE

2.1. ZIELSETZUNG UND METHODIK DER BESTANDSAUFNAHME

2.1.1. Problemstellung	11
2.1.2. Abgrenzung des Forschungsfeldes	16
2.1.2.1. Die Kriterien	16
2.1.2.2. Komplexes Informationsproblem	16
2.1.2.3. Effizienz als abhängige Variable	17
2.1.2.4. Informationsverhalten als unabhängige Variable	18
2.1.2.5. Verfügbarkeit der Studien	22
2.1.2.6. Keine Vorauswahl aufgrund methodischer Standards	25
2.1.3. Beurteilung der empirischen Studien	26

2.2. INFORMATION-SYSTEM UND EFFIZIENZ

2.2.1. Überblick	32
2.2.2. Angebotsmenge und Effizienz	32
2.2.2.1. Die Hypothesen	32
2.2.2.2. Effizienz betrieblicher Entscheidungen – jenseits einer kritischen Mindestmenge unabhängig von der Angebotsmenge	34
2.2.2.3. Effizienz von Kaufentscheidungen: divergierende Befunde zum Einfluß von Alternativen- und Attributmengen	41
2.2.2.4. Wirkungen der Angebotsmenge – ein Zwischenergebnis	50

2.2.3. Aggregationsgrad und Effizienz	51
2.2.3.1. Die Hypothese	51
2.2.3.2. Die Ergebnisse: Gewisse Anzeichen für eine Bestätigung des umgekehrt u-förmigen Verlaufes	53
2.2.3.3. Zusammenfassende Würdigung	58
2.2.4. Angebotsmedium und Effizienz	58
2.2.5. Graphische Unterstützung und Effizienz	62
2.2.5.1. Die Hypothesen	62
2.2.5.2. "Tabelle vs. Graphik": Keine praktisch bedeutsamen Unterschiede	64
2.2.5.3. Zuverlässigere Diagnose von Unternehmensentwicklungen mit "Schematischen Gesichtern"?	70
2.2.5.4. Unterstützung der Bilanzanalyse durch Baumstrukturen	74
2.2.5.5. Erste Anzeichen für positive Wirkungen mehrfarbiger Informationsangebote	74
2.2.6. Effizienzsteigerung bei relevantem Feedback	75
2.2.7. Formale Entscheidungshilfen: noch wirksamer bei angemessenem Informationsverhalten	78

2.3. PERSÖNLICHKEIT UND EFFIZIENZ BETRIEBLICHER ENTSCHEIDUNGSPROZESSE

2.3.1. Problemstellung	83
2.3.2. Die Grundtendenz: Weder "Kognitive Stile" noch "Test-Intelligenz" führen generell zu besseren Leistungen	85
2.3.3. Einflüsse Kognitiver Stile	89
2.3.3.1. Perzeptionsbezogene Stile: Effizienzerklärung bei stilgerechten Aufgaben	90
2.3.3.2. Myers-Briggs Indikatoren: Keine praktisch bedeutsamen Einflüsse eines besonders häufig empfohlenen Stilkonzeptes	95
2.3.3.3 "Analytisch" vs. "Heuristisch": keine praktisch bedeutsamen Einflüsse selbst-gestrickter Meßkonzepte	98
2.3.3.4. Sonstige Stile: Tendenzielle Einflüsse, die weiterer Prüfung bedürfen	100
2.3.4. Einflüsse von Test-Intelligenz und quantitativen Fähigkeiten	104
2.3.4.1. Test-Intelligenz: Effizienzeinflüsse bei testgerechten Aufgaben	104
2.3.4.2. Quantitative Fähigkeiten: nur selten beachtliche Einflüsse auf die Effizienz	111
2.3.5. Konsequenzen: Plädoyer für eine differenzierte Forschung	111

2.4. INFORMATIONSVERHALTEN UND EFFIZIENZ

2.4.1. Überblick	117
2.4.1.1. Abgrenzung der untersuchten Studien	117
2.4.1.2. Die Hypothesen	121
2.4.1.3. Zur Methodik der Bestandsaufnahme	125
2.4.2. Ergebnisse der Feldstudien	126
2.4.2.1. Methodische Beurteilung der Studien	126
2.4.2.2. Befunde zur Baustein-Erklärung: allenfalls schwache Belege für die anspruchslose Hypothesenvariante	130
2.4.2.3. Deutliche Hinweise auf die Gültigkeit der Verarbeitungs-Hypothese	131
2.4.2.4. Bestätigung der Steuerungs-Hypothese	132
2.4.2.5. Zusammenfassende Würdigung der Ergebnisse	133
2.4.3. Ergebnisse der experimentellen Studien mit computergetstützten Managementspielen	134
2.4.3.1. Informationsnachfrage und Effizienz: zahlreiche Belege für einen positiven Zusammenhang	134
2.4.3.2. Hypothesenbildung, Analysen und Planungen: differenzierende Belege für die Erkenntnis- und Verarbeitungs-Erklärung	143
2.4.4. Ergebnisse der Beurteilungs- und Prognosestudien	150
2.4.4.1. Befunde zur Baustein-Hypothese: Bestätigung der anspruchsvollen, Falsifikation der anspruchslosen Variante	150
2.4.4.2. Erkenntnis- und Verarbeitung-Hypothese: positive Evidenz weniger Prüfungen	154
2.4.5. Zusammenfassende Würdigung	155

3. TEIL: MESSWERTE DER EFFIZIENZ VON BILANZANALYSEN – THEORETISCHES KONZEPT UND EMPIRISCHE VALIDIERUNG –	110
3.1. THEORETISCHE GRUNDLEGUNG	156
3.1.1. Ziele der Effizienzmessung	156
3.1.2. Dimensionen der Effizienzmessung	160
3.1.2.1. Überblick	161
3.1.2.2. Humaneffizienz	161
3.1.2.3. Kosten der Analyse	162
3.1.2.4. Nutzen der Analyse	164
3.1.3. Vergleich mit Meßkonzepten anderer Untersuchungen	168
3.1.3.1. Überblick	168
3.1.3.2. Ressourcenverbrauch	169
3.1.3.3. Subjektive Leistungseinschätzung	170
3.1.3.4. Ökonomischer Erfolg	170
3.1.3.5. Prognosegüte	171
3.1.3.6. Angemessenheit von Beurteilungen	172
3.2. OPERATIONALISIERUNG DER ANDEREN EFFIZIENZDIMENSIONEN	200
3.2.1. Die Operationalisierung der Urteilsrelevanz	202
3.2.2. Die Operationalisierung der Urteilsüberzeugungsrelevanz	202
3.2.3. Die Operationalisierung der Kostenrelevanz	202
3.2.4. Die Operationalisierung der Nutzenrelevanz	202
3.2.5. Die Operationalisierung der Vergleichsrelevanz	202
3.2.6. Die Operationalisierung der Ressourcenverbrauchsrelevanz	202
3.2.7. Die Operationalisierung der Subjektiven Leistungseinschätzungsrelevanz	202
3.2.8. Die Operationalisierung der Ökonomischen Erfolgsrelevanz	202
3.2.9. Die Operationalisierung der Prognosegüterrelevanz	202
3.2.10. Die Operationalisierung der Angemessenheitsrelevanz	202
3.3. OPERATIONALISIERUNG DER QUALITÄT DER URSACHENANALYSE	202
3.3.1. Die Operationalisierung der Qualität der Ursachenanalyse	202
3.3.2. Die Operationalisierung der Qualität der entwicklungsfördernden Folgen	202
3.3.3. Die Operationalisierung der Humaneffizienz	202

3.2. OPERATIONALISIERUNG DER TREFFSICHERHEIT DER LAGEBEURTEILUNG	
3.2.1. Das Meßziel	193
3.2.2. Die Bestimmung der Urteilsbereiche	200
3.2.3. Die Bestimmung der Relevanzmaße	202
3.2.3.1. Literarische Relevanz	202
3.2.3.2. Statistische Relevanz	211
3.2.3.3. Fallspezifische Relevanz	222
3.2.4. Die Prüfung der qualitativen Anforderungen an die Urteilsbildung	225
3.2.5. Die Meßwerte zur Treffsicherheit der Lagebeurteilung	233
3.2.5.1. Die Bestimmung des Hauptmeßwertes	235
3.2.5.2. Differenzierungen des Hauptmeßwertes	245
3.3. OPERATIONALISIERUNG DER ANDEREN EFFIZIENZDIMENSIONEN UND DER BEURTEILUNGSTENDENZ	
3.3.1. Die Operationalisierung der Qualität der Ursachenanalyse	248
3.3.2. Die Operationalisierung der Qualität der entwickelten Fragen	251
3.3.3. Die Operationalisierung der Humaneffizienz	255

3.3.4. Die Operationalisierung der Beurteilungstendenz	256
3.3.4.1. Beurteilungstendenz des Gesamturteils	256
3.3.4.2. Beurteilungstendenz der Teilurteile	257
3.4. EMPIRISCHE ANALYSE UND VALIDIERUNG DES MESSKONZEPTE	290
3.4.1. Überblick	259
3.4.2. Beschreibung des Urteilsverhaltens und Analyse der Einflüsse von Versuchssituation und Versuchspopulation	261
3.4.2.1. Beschreibung des Ausgangsmaterials	261
3.4.2.2. Vergleich des Urteilsverhaltens bei den experimentellen Fällen und beim Klausurfall	266
3.4.2.3. Vergleich des Urteilsverhaltens von studentischen Analytikern und Wirtschaftspraktikern	268
3.4.3. Entwicklung der Effizienzmeßwerte über die einzelnen Fälle	272
3.4.3.1. Entwicklung der Meßwerte zur Treffsicherheit der Lagebeurteilung	272
3.4.3.2. Entwicklung der anderen Effizienzdimensionen und der Beurteilungstendenz	276
3.4.4. Auswertungen zur Konsistenz interindividueller Leistungsunterschiede bei verschiedenen Fällen	278
3.4.5. Pfadanalyse zur Konstruktvalidierung des Effizienzkonzeptes	282

4. TEIL: EMPIRISCHE UNTERSUCHUNG AM BEISPIEL VON BILANZANALYSEN	301
4.1. THEORETISCHER BEZUGSRAHMEN DER EMPIRISCHEN UNTERSUCHUNG	301
4.1.1. Das Grundmodell	287
4.1.2. Die Elemente des Informationsverhaltens	290
4.1.2.1. Die Informationsaktivitäten	290
4.1.2.2. Die Informationsobjekte	292
4.1.2.3. Zusammenfassung des Meßkonzeptes	293
4.1.3. Die Dimensionen des Informationsverhaltens	294
4.1.3.1. Die strategischen Dimensionen	295
4.1.3.2. Die taktischen Dimensionen	296
4.1.3.3. Die operativen Dimensionen	297
4.1.4. Hypothesen zum Zusammenhang der Dimensionen	298
4.1.4.1. Abgrenzung	298
4.1.4.2. Zusammenhänge zwischen den strategischen Dimensionen	299
4.1.4.3. Zusammenhänge zwischen den taktischen Dimensionen	299
4.1.4.4. Zusammenhänge zwischen den operativen Dimensionen	300

4.1.5. Hypothesen zur Effizienzwirkung des Informationsverhaltens	301
4.1.5.1. Effizienzwirkungen der Informationsbeschaffung	301
4.1.5.2. Effizienzwirkungen der Übertragung und Speicherung	303
4.1.5.3. Effizienzwirkungen der Kennzahlenbildung	304
4.1.5.4. Effizienzwirkungen des Aggregationsgrades	306
4.1.5.5. Effizienzwirkungen der Vergleichstechnik	308
4.1.5.6. Effizienzwirkungen der Abfolge der Verrichtungen	308
4.1.5.7. Effizienzwirkungen der Abfolge der Objekte	310
4.1.6. Die Einflüsse der Rahmenvariablen	318
4.1.6.1. Die Einflüsse der Präsentationsform	319
4.1.6.2. Die Einflüsse der Erfahrung	322
4.1.7. Zusammenfassung der Hypothesen	324
4.2. METHODIK UND DURCHFÜHRUNG DER EMPIRISCHEN UNTERSUCHUNG	327
4.2.1. Überblick	326
4.2.2. Das Untersuchungsobjekt	328
4.2.3. Die Versuchspersonen	333

4.2.4. Design und Ablauf der Experimente	341
4.2.5. Die Validität der Experimente	348
4.2.5.1. Problemgehalt und Überblick	348
4.2.5.2. Die Prüfung von Gruppeneffekten	353
4.2.5.3. Die Prüfung von Arbeitsplatzeffekten	354
4.2.5.4. Die Prüfung von Ermüdungseffekten	355
4.2.6. Die Messung des Informationsverhaltens	357
4.2.6.1. Die Erhebungsmethoden	357
4.2.6.2. Die operationalen Definitionen der Informationsaktivitäten	357
4.2.6.3. Die Prüfung der Reliabilität der Informationsaktivitäten	359
4.2.6.4. Die operationalen Definitionen der Dimensionen des Informationsverhaltens	359
4.2.6.5. Die Prüfung der Validität der Dimensionen des Informationsverhaltens	361
4.2.7. Die statistischen Auswertungsverfahren	367
4.2.7.1. Das Auswertungsprogramm	367
4.2.7.2. Die Prüfung der Verfahrensvoraussetzungen	372
4.2.7.3. Die Zurückweisungs- und Beurteilungsregeln	376

4.3.1.3. Die Präanalyse der Effizienzwirkungen	401
4.3. DIE ERGEBNISSE DER EMPIRISCHEN UNTERSUCHUNG	401
4.3.1. Die interpersonelle Analyse	379
4.3.1.1. Interpersonelle Effizienzunterschiede: Der Vergleich guter und schlechter Analytiker	379
4.3.1.1.1. Der Ausgangsbefund: Sind alle Effizienzhypothesen zum Informationsverhalten falsifiziert?	379
4.3.1.1.2. Prüfung der Variation des Informationsverhaltens: keine Einwände gegen die Tests	382
4.3.1.1.3. Erste Differenzierung: Falleinfluß und Lernfortschritt	384
4.3.1.1.4. Zweite Differenzierung: Effekte unterschiedlicher Analysefelder	387
4.3.1.1.5. Zusammenfassung der direkten Effekte: Informationsverdichtung und Vorbildung sind beachtlich	390
4.3.1.2. Interpersonelle Unterschiede im Informationsverhalten	393
4.3.1.2.1. Die Einflüsse der Rahmenvariablen: deutliche Effekte der Vorbildung	393
4.3.1.2.2. Die Dimensionen des Informationsverhaltens: unabhängige Aktionsparameter	397
4.3.1.3. Differenzierung nach Populationsmerkmalen	437
4.3.2. Zusammenfassende Würdigung der intertemporalen Analyse	439

Gliederung	XII
4.3.1.3. Die Pfadanalyse der Effizienzwirkungen	401
4.3.1.3.1. Zur Methodik der Auswertungen	401
4.3.1.3.2. Treffsicherheit der Gesamtlage: dysfunktionale indirekte Effekte der Vorbildung	403
4.3.1.3.3. Unterschiedliche Analysefelder: differentielle Wirkungen von Präsentationsform und Vorbildung	411
4.3.1.4. Zusammenfassende Würdigung der interpersonellen Analyse	416
4.3.2. Die intertemporale Analyse: das Lernverhalten der Analytiker	419
4.3.2.1. Zur Methodik der intertemporalen Analyse	419
4.3.2.2. Der Ausgangsbefund: Erneute Falsifikation der umgekehrt u-förmigen Verlaufshypothese	421
4.3.2.3. Die Entwicklung des Informationsverhaltens: unterschiedlich ausgeprägt bei den einzelnen Dimensionen	423
4.3.2.4. Differenzierung nach Analysefeldern: engerer Zusammenhang bei der Erfolgslage	425
4.3.2.5. Pfadanalyse der Effizienzwirkungen: Die Veränderung des Analysekonzeptes als Ursache des Lernfortschritts	429
4.3.2.5.1. Gemeinsame Analyse aller Informationsprozesse: beachtliche Effekte der Objektgliederung	429
4.3.2.5.2. Differenzierung nach Präsentationsformen: Bestätigung der Verhaltensmuster	434
4.3.2.5.3. Differenzierung nach Populationen: Erneute Bestätigung der Verhaltensmuster	437
4.3.2.6. Zusammenfassende Würdigung der intertemporalen Analyse	439

1. TEIL: PROBLEMSTELLUNG

In einer dynamischen und komplexen Umwelt, die das Management vor immer anspruchsvollere Probleme stellt, kommt der Information, dem dynamischen Element der Entscheidung, eine zentrale Rolle zu. Es ist daher nicht verwunderlich, daß eine Betriebswirtschaftslehre, die sich in zunehmendem Maße dem entscheidungsorientierten Wissenschaftsprogramm¹⁾ verschrieben hat, die Zusammenhänge zwischen Information und Entscheidungsqualität intensiv erforscht:

- * Substantiell ausgerichtete Beiträge der **Speziellen Betriebswirtschaftslehren** widmen sich der Bestimmung des **Informationsbedarfs** der jeweils untersuchten Probleme.²⁾ Eine besondere Stellung nimmt die **Unternehmensrechnung** ein, die sich auch den Fragen der zweckmäßigen **Abbildung** betriebswirtschaftlicher Güterströme und Dispositionen zuwendet.³⁾
- * **Informationsökonomische Untersuchungen** treffen grundlegende Aussagen über **Kosten und Nutzen** von Informationen und Informationsaktivitäten.⁴⁾ Sie stützen sich dabei häufig auf vereinfachende Prämissen, deren empirische Prüfung noch aussteht.
- * Untersuchungen der **Wirtschaftsinformatik** befassen sich mit dem Einsatz geeigneter **Technologien** zur Gewinnung, Verarbeitung, Speicherung und Präsentation von Informationen im Rahmen betriebswirtschaftlicher **Management-Information- und Entscheidungs-Unterstützungs-Systeme** (im folgenden kurz als "Informationssystem" bezeichnet).
- * **Verhaltensorientierte Ansätze** stellen die Merkmale der Entscheidungsträger und deren (tatsächliches) **Informationsverhalten** in den Vordergrund. Einflüsse auf die Effizienz der Entscheidung werden vergleichsweise selten erforscht.

Alle vier Wege sind notwendig. Erst zusammengekommen sind sie hinreichend, eine nachhaltige Verbesserung betrieblicher Entscheidungen zu bewirken. Der einzelne Forscher muß jedoch eine begründete Auswahl treffen, welchen Weg er beschreiten möchte.

Unser besonderes Forschungsinteresse gilt dem letzten Bereich. Wir untersuchen die Wirkungen des Informationsverhaltens auf die Effizienz betrieblicher Entscheidungen. Dies ist eine wichtige und vor allem in der deutschen betriebswirtschaftlichen Forschung noch stark vernachlässigte Problemstellung. Der Nachweis für diese These wird im Verlauf der Arbeit durch eine systematische Bestandsaufnahme der empirischen Forschungsergebnisse und eigene empirische Untersuchung erbracht werden. Einige programmatische Überlegungen zur Relevanz des Themas sollen vorangestellt werden.

Vereinfacht gesprochen können Verbesserungen von Entscheidungen an drei Faktoren ansetzen:⁵⁾

- a) dem **Problem**, das gelöst werden soll,
- b) dem **Informationssystem**, das die Problemlösung unterstützt,
- c) dem **Menschen**, der das Problem lösen soll.

Bei allen drei Faktoren ist eine verhaltensorientierte Analyse einerseits zweckmäßig, andererseits bisher nicht konsequent genutzt worden.

a) Ansatzpunkt "Problem":

Eine bessere Entscheidung wird allgemein erwartet, wenn Probleme zweckmäßig abgegrenzt und strukturiert werden. Entscheidungshilfen versprechen dann positive Wirkungen, wenn ihre Anwendungsprämissen erfüllt sind. Ein häufig beschrittener Forschungsweg besteht deshalb darin, die anfallenden Aufgaben idealtypisch zu beschreiben und für diese Idealtypen Gestaltungsempfehlungen abzuleiten. Eine Schwachstelle dieses Ansatzes stellt die Beurteilung der Ähnlichkeit von idealtypischer Aufgabe und realem Problem und die Überprüfung (der Gewichtigkeit) der gesetzten Prämissen dar. Vernachlässigt man diese Fragen, so kann es insbesondere zu einem sogenannten "Fehler dritter Art" kommen: es werden Empfehlungen für Probleme abgegeben, die sich realiter nicht stellen.⁶⁾

Aus diesem Grunde ist es sinnvoll, ergänzend einen zweiten Weg zu beschreiben.⁷⁾ Er geht vom Menschen aus und erforscht, wie Entscheidungsträger ihre Probleme definieren und strukturieren und welche Wirkungen dies auf die Effizienz ihrer Entscheidungen hat. Aus der Gegenüberstellung von idealtypisch beschriebener Aufgabe und subjektiver Problemdefinition lassen sich Rückschlüsse ziehen auf beiderseitige Defizite und Barrieren, die sich den Gestaltungsempfehlungen entgegenstellen.

In der vorliegenden Arbeit wird der seltener beschrittene zweite Weg gewählt. An Hand des komplexen Beurteilungsproblems "Bilanzanalyse" wird empirisch geprüft, ob die Analysekonzepte bilanzanalytischer Monographien dem tatsächlichen Informationsverhalten entsprechen und welche Effizienzwirkungen mögliche Diskrepanzen haben. Wir glauben, daß sich dieser Bereich für eine Prüfung besonders gut eignet, weil für dieses wichtige und häufig auftretende Problem einerseits zahlreiche Techniken empfohlen werden, andererseits aber noch nicht empirisch geprüft wird, ob diese Techniken tatsächlich zu besseren Analyseergebnissen führen.⁸⁾

b) Ansatzpunkt "Informationssystem":

Betrachtet man die stürmische Ausbreitung von "personal computer", die zunehmende Vernetzung von Kommunikationssystemen, den wachsenden Einsatz graphischer Datenverarbeitungstechniken und die zunehmende Integration betrieblicher Informationsströme, dann kann man den Eindruck gewinnen, daß der Schlüssel zur besseren und schnelleren Lösung dieser Probleme vor allem in leistungsfähigeren Informationssystemen gesehen wird.

Doch vor übertriebener Technologie-Euphorie haben Kritiker bereits bei der Einführung von Großrechnern gewarnt: Die bloße Verfügbarkeit großer Informationsmengen und die superschnelle Ausführung arithmetischer und logischer Operationen sichern noch keine Effizienzsteigerung. Es käme erstens darauf an, "relevante" Informationen zu speichern und eine **problemgerechte** Verknüpfung und Verdichtung von Informationen vorzunehmen. Zweitens müßten die Programm- und Datenbanken **subjektgerecht** konstruiert werden, damit ein problemgerechtes Informationsangebot auch genutzt würde.⁹⁾

Mißerfolge belegten, daß diese Kritik berechtigt war. Sie führten zum Schlagwort Management-"Mis"-Informationssystem, zum berüchtigten "Zahlenfriedhof" und zur These des "Verdurstens im Informationsüberfluß". Sie bewirkten allerdings auch, daß man zukünftige Benutzer stärker an der Entwicklung und Implementierung von computergestützten Informationssystemen beteiligte, leistungsfähigere Sprachen entwickelte, die Möglichkeiten der Dialogführung mit dem Informationssystem verbesserte und Entscheidungshilfen auch auf Akzeptanz und Praktikabilität hin überprüfte.

Es ist jedoch zu bezweifeln, ob damit eine grundlegende Korrektur des Verständnisses betrieblicher Informationsprobleme herbeigeführt wurde. Betrachtet man Termini wie benutzer-"freundliche" Software, ansprechende Benutzer-"Oberfläche" oder gute "Bediener"-Führung, dann erkennt man rasch, daß derartige Vorkehrungen vor allem die erst von den maschinellen Systemen errichteten **Interaktionsbarrieren** der Mensch-Maschine-Kommunikation überwinden sollen.

Doch die flexiblen, modernen Lösungen mit relationalen Datenbankumgebungen, Endbenutzersprachen und neuartigen Ein- und Ausgabemedien richten sich in immer stärkerem Maße auf einmalige, schlecht-strukturierte, **komplexe** Probleme. Sie geben dem Menschen nicht nur mehr Gestaltungsfreiheiten, sie machen ihn auch verantwortlich für eine größere Bandbreite effizienter und ineffizienter Lösungen. Aus diesem Grunde nimmt die Bedeutung des "Faktors" Mensch im Mensch-Maschine-System nicht etwa ab, sondern zu. Deshalb darf man sich auch nicht auf "Schnittstellen"-Probleme der Mensch-Maschine-Kommunikation beschränken. Man muß tiefer ansetzen, muß auch die **Komplexitätsbarriere** des Problemlösers berücksichtigen.¹⁰⁾

Die darauf gerichtete Analyse muß vom menschlichen Informationsverhalten ausgehen. Man muß systematisch erforschen, welche Schwächen, welche **Insuffizienzen**, das menschliche Informationsverhalten aufweist, um den Bedarf an Servo-Mechanismen zu erkennen, die das Informationssystem bereitstellen sollte, um sie zu beheben. Andererseits kann man aus den **Stärken** guter Problemlöser lernen, welche Informationen bereitzustellen sind und wie man sie verknüpft.¹¹⁾ Das strapazierte Schlagwort "Expertensystem" belegt die

Aktualität dieser Anstrengungen.¹²⁾

Eine solche Forschung ist aufwendig, aber notwendig: Wie wir in unserer kritischen Bestandsaufnahme der vorliegenden Effizienzprüfungen belegen werden, konnte weder bei klassischen Gestaltungsvariablen wie "Menge" und "Aggregationsgrad" angebotener Daten, noch bei späteren Vergleichen von "Bildschirm vs. Liste", noch in den neuesten Untersuchungen zur Effizienzwirkung farbiger Graphiken ein starker Einfluß durchgängig belegt werden. Die meisten Studien zeigen vielmehr **keine** signifikanten Unterschiede.¹³⁾ Das Problematische an diesen Studien ist, daß man darauf verzichtet hat, das tatsächliche Informationsverhalten zu messen und zur Effizienz in Beziehung zu setzen. Man weiß also gar nicht, **warum** die Studien kein signifikantes Ergebnis zeigen. Deshalb fordern wir eine theoriegeleitete pfadanalytische Untersuchung der behaupteten direkten und indirekten Effizienzwirkungen, in der das tatsächliche prozeßbegleitend gemessene Informationsverhalten differenziert berücksichtigt wird: **Wer wirksame Informationssysteme schaffen will, der sollte auch differenziert prüfen, wie sie auf das Informationsverhalten des Anwenders und dessen Effizienz wirken.** Unsere eigene empirische Studie stellt sich diesem Anspruch.

c) Ansatzpunkt "Mensch":

Maßnahmen zur Verbesserung betrieblicher Entscheidungen können nach wie vor und nicht zuletzt direkt am Menschen ansetzen. Geeignete Sozialtechnologien können die Diffusion innovativer Konzepte beschleunigen und die Akzeptanz fortschrittlicher Problemlösungen erhöhen. Doch mit einer reinen Werbung für "gute" Ideen ist es nicht getan. Man muß die Benutzer auch im Umgang mit neuartigen Informationssystemen schulen, sie frühzeitig an deren Entwicklung beteiligen, um subjektgerechte Lösungen zu erreichen, mit denen sich die Anwender identifizieren. Soweit eine Anpassung des Menschen nicht möglich oder nicht zumutbar erscheint, stellt sich als weitere Option die Handlungsmöglichkeit, Mitarbeiter zu rekrutieren, die für die veränderte Aufgabe besser geeignet sind.

Zu diesen Handlungsoptionen gibt es quantitativ und qualitativ unterschiedlich ausgeprägte empirische Studien.

- * Vergleichsweise intensiv untersucht werden Erscheinungsbild, Verlauf und Wirkungen des sog. **"Implementierungsprozesses"**.¹⁴⁾ Die Ergebnisse dieser Forschungsrichtung werden von uns nicht weiter verfolgt, da wir uns auf individuelle Entscheidungen konzentrieren.¹⁵⁾
- * In der anglo-amerikanischen Forschung sind empirische Studien, in denen die Einflüsse unterschiedlicher **"Kognitiver Stile"** auf Informationsverhalten und Effizienz komplexer Problemlösungsprozesse untersucht werden, weit verbreitet.¹⁶⁾ In diesen Studien werden häufig bestimmte Merkmale des Informationssystems variiert, um die **Kontingenzhypothese** zu überprüfen, daß bestimmte Typen von Informationssystemen bestimmten Typen von Entscheidungsträgern entsprechen sollten. Wie unsere Bestandsaufnahme zeigen wird, ist die empirische Evidenz ziemlich dürftig: In der großen Mehrheit der Studien zeigen sich weder signifikante Haupteffekte noch signifikante Interaktionen mit Merkmalen des Informationssystems auf ökonomische Effizienzgrößen. Ein ähnliches Ergebnis ergibt sich für die zahlreichen neueren Studien aus der deutschsprachigen psychologischen Problemlöseforschung: Die auf verschiedene Weisen gemessene **"Test"-Intelligenz** steht in keiner systematischen Beziehung zu ökonomischen Effizienzgrößen in computergestützten Simulationen von komplexen Managemententscheidungen.
- * Eine weitaus geringere Zahl empirischer Untersuchungen setzt das **prozeßbegleitend gemessene Informationsverhalten** zur Effizienz der individuellen Entscheidung in Beziehung. Obwohl die einzelnen Studien sehr unterschiedlich angelegt sind, zeigen sie mehrheitlich doch einen deutlichen Zusammenhang beider Größen.
- * Noch seltener wird erforscht, wie sich das Informationsverhalten im Zeitablauf verändert und in welcher Beziehung dieser **Wandel** des Informationsverhaltens zur Veränderung der Effizienz steht.

Insgesamt muß der Stand der sog. "verhaltensorientierten" Forschung als außerordentlich unbefriedigend angesehen werden.¹⁷⁾ Der Versuch, die Effizienz des Verhaltens bei komplexen betriebswirtschaftlichen Entscheidungen aufgrund von Persönlichkeitsmerkmalen, die in gängigen Testinventaren bereitgestellt oder von Betriebswirten (weiter-) entwickelt werden, ist als gescheitert anzusehen. Eine Rekrutierungsentscheidung oder eine subjektangepaßte Auslegung von Informationssystemen kann auf Basis der vorliegenden Untersuchungen nicht getroffen werden.

Die wenigen Studien, die ein prozeßbegleitend gemessenes Informationsverhalten zur Effizienz in Beziehung setzen, divergieren stark im methodischen Ansatz. Sie gehen in der Regel heuristisch-explorativ vor und lassen die Frage weitgehend offen, auf welchen "Hintergrundvariablen" die beobachteten Effizienzwirkungen letztlich beruhen. Die große Mehrheit der Studien konzentriert sich auf eine einmalige Problemlösung und bleibt den Nachweis schuldig, daß es sich um reproduzierbare Zusammenhänge handelt. Außerdem beschränkt sich die Messung des Informationsverhaltens häufig auf die Informationsbeschaffung, namentlich auf die Menge an beschafften Informationen. Es lassen sich daher nur sehr allgemein gehaltene praxeologische Konsequenzen ableiten, und auch diese sind mit einer erheblichen Unsicherheit belastet.

Zur Frage, wie sich das Lernverhalten vollzieht und auf die Effizienz komplexer betriebswirtschaftlicher Entscheidungen auswirkt, ist uns keine differenzierte empirische Untersuchung bekannt. Es gibt daher auch keine Gestaltungshinweise auf entsprechende Schulungsprogramme oder Prognosehilfen für eine dynamische Abschätzung der Effizienz von Informationssystemen.

Diese Aussagen stehen nicht im Widerspruch zu der großen Menge empirischer Studien über das menschliche Informationsverhalten.¹⁸⁾ Doch leider gilt auch hier der Satz, daß man in der Informationsschwemme verdurstet. Viele Studien scheiden für unsere Zwecke aus, weil sie ein zu simples Problem betrachten, sich auf nachträglich erinnertes oder hypothetisch behauptetes Informationsverhalten beschränken, oder weil sie keinerlei Beziehungen zur Effizienz herstellen. Aus diesem Grunde ist ein **erhebliches Defizit** an betriebswirtschaftlich relevanter Forschung zum Informationsverhalten festzu-

stellen. Das Defizit betrifft nicht die Quantität, sondern die Art und die Qualität der für unsere Fragestellung erforderlichen empirischen Studien.

Aus diesen Überlegungen ergibt sich das Forschungsprogramm der vorliegenden Arbeit:

1. Wir wollen im **zweiten Teil** der Arbeit eine **kritische Bestandsaufnahme der empirischen Studien** zum Zusammenhang von Informationssystem, Informationsverhalten und Effizienz bei individuellen komplexen betriebswirtschaftlichen Entscheidungsproblemen vornehmen. Diese Bestandsaufnahme soll über die wichtigsten Erklärungsansätze, Methoden, Ergebnisse, aber auch über die Defizite der weit verstreuten Untersuchungen informieren. Sie soll dazu beitragen, Doppelarbeiten zu vermeiden und Defizite zu überwinden. Sie soll dafür werben, sich mit diesen in der deutschen betriebswirtschaftlichen Forschung stark vernachlässigten Problemen intensiver zu befassen.

Die weitere Arbeit richtet sich auf die theoretische und empirische Analyse des komplexen betriebswirtschaftlichen **Beurteilungsproblems Bilanzanalyse**. Eine Konzentration auf einen bestimmten Problemtyp ist unseres Erachtens notwendig, wenn man Effizienzaussagen ableiten und prüfen will, weil die Effizienzbestimmung von der Art der Aufgabe abhängt. Außerdem ermöglicht sie eine differenziertere Analyse und eine Ableitung von Aussagen höherer Bestimmtheit, was für die praktische Anwendung wichtig ist. Schließlich kann man nur bei einem konkreten Problem das **Zusammenwirken** der Faktoren Informationssystem, Informationsverhalten und Effizienz systematisch und differenziert erforschen.

Unsere Aussagen können vermutlich auch bei anderen betriebswirtschaftlichen Beurteilungsproblemen aus den Bereichen Planung, Controlling, Revision und Marktforschung entsprechend modifiziert Geltung beanspruchen. Unser beschränktes Wissen über diese Aufgaben und der unbefriedigende augenblickliche Stand der verhaltensorientierten Forschung lassen es nicht zweckmäßig erscheinen, diesen Anspruch zu erheben. Wir folgen lieber der auf langjähriger Erfahrung basierenden Empfehlung Wittes:

"... zunächst konkrete Einzelprojekte zu bearbeiten und auf deren festem Boden aufzubauen, um allgemeinere theoretische Konstrukte anzustreben."¹⁹⁾

Denn:

"Der umgekehrte Weg, zunächst weit ausgreifende Entwürfe zu formulieren und auf dem Wege der Deduktion aus zunächst ungesicherten Vermutungen zu konkreten Einzelfragen vorzudringen, hat sich forschungsmethodisch nicht bewährt."²⁰⁾

2. Im **dritten Teil** dieser Schrift wollen wir ein theoretisches Konzept zur **Messung der Effizienz** von Bilanzanalysen vorstellen und empirisch validieren. Obwohl es sich dabei "nur" um die Operationalisierung der abhängigen Variablen unserer empirischen Studie handelt, scheint es uns angesichts der Bedeutung und Komplexität der Meßprozedur sinnvoll, diesen Untersuchungsschritt in einen eigenen Teil der Schrift auszugliedern. Damit soll zugleich eine Entlastung des Lesers ermöglicht werden: Die Arbeit ist so konzipiert, daß der Leser diesen Teil zunächst überspringen und bei Bedarf nachlesen kann.

3. Im **vierten Teil**, dem Hauptteil dieser Arbeit, wird ein **theoretischer Bezugsrahmen** vorgelegt, der die verschiedenen Erklärungsskizzen integriert und zu überprüfbaren Effizienzhypothesen verdichtet. Außerdem werden Hypothesen zur Entwicklung des Informationsverhaltens abgeleitet. Anschließend wird das **experimentelle Design**, auf das sich diese Studie stützt, beschrieben. Einflüsse, die die externe, interne, statistische und Konstruktvalidität beeinträchtigen können, werden diskutiert und empirisch überprüft. Schließlich werden die **Befunde** präsentiert, wobei zwischen einer interpersonellen und einer intertemporalen Analyse differenziert wird.

Die vorliegende Arbeit wurde durch ein Habilitationsstipendium der Deutschen Forschungsgemeinschaft gefördert. Die ausgewerteten Daten der Experimentalserien Gießen 82 und Kiel 82 wurden im Rahmen des von der Forschungsgemeinschaft finanzierten Projektes "Informationsgleichgewicht" erhoben. Die Erhebung der Daten der Serie Kiel 84 wurde von der Konrad-Adenauer-Stiftung und vom Institut für Betriebswirtschaftslehre der Christian-Albrechts-Universität zu Kiel finanziert. An diesen Experimenten waren neben dem Projektleiter, Professor Dr. Jürgen Hauschildt und dem Verfasser weitere Mitarbeiter des Institutes für Betriebswirtschaftslehre beteiligt. Diesen Institutionen und Personen, namentlich den Herren Prof. Dr. Jürgen Hauschildt, Dipl.-Wirt. Ing., Dipl.-Kfm. Wolfgang Fink, Prof. Dr. Peter Knorr, Dr. Harald Krehl und Dipl.-Kfm. Knut Petersen, danke ich herzlich.²¹⁾

Eigenschaften effizientsteigern²²⁾

* als Indikator zur Steuerung bedarfsgerechter Informationsanforderungen ("Steuerungs-Erklärung")

* als Signal freier zeitlicher und kognitiver Kapazität sowie vorhandenem Arbeitswillens ("Kapazitäts-Erklärung")

* als Indikator für Problemlösungstendenz ("Erkenntnis-Erklärung").

In unseren Untersuchungen konnten diese Thesen nicht empirisch bestätigt werden. Es ergaben sich sowohl theoretisch begründete Indikatoren als auch konträre empirische Ergebnisse:

* Informationsnachfrage kann durch vorhandenes (experientielles) Wissen substituiert werden.²³⁾

* hohe Informationsnachfrage ist auch Ausdruck von Unsicherheit, Zielkonflikten²⁴⁾ und ungenügender Kenntnis des Informationsangebots.²⁵⁾

* selbst-induzierte, dysfunktional wirksame Informationsüberflutung kann nicht ausgeschlossen werden.²⁶⁾

2. TEIL: DETERMINANTEN DER EFFIZIENZ INDIVIDUELLER ENTSCHEIDUNGEN EINE SYSTEMATISCHE BESTANDSAUFNAHME EMPIRISCHER FORSCHUNGSERGEBNISSE

2.1. ZIELSETZUNG UND METHODIK DER BESTANDSAUFNAHME

2.1.1. Problemstellung

Das fokale Interesse dieser Schrift gilt dem Zusammenhang von Informationsverhalten und Effizienz. Die empirische Analyse dieses Zusammenhanges hat in der deutschen betriebswirtschaftlichen Forschung bereits Tradition. Sie wurde begründet durch Witte, der in der "Vitalisierung der unvollkommenen Informations-Nachfrage" den Schlüssel zur Effizienzsteigerung sieht.¹⁾ Nach seinem Erklärungsansatz wirkt eine hohe Informationsnachfrage in drei Eigenschaften effizienzsteigernd:²⁾³⁾

- * als Instrument zur Steuerung bedarfsgerechter Informationsversorgung ("Steuerungs-Erklärung")
- * als Signal freier zeitlicher und kognitiver Kapazität sowie vorhandenen Arbeitswillens ("Kapazitäts-Erklärung")
- * als Indikator für Problemeinsicht ("Erkenntnis-Erklärung").

In neueren Untersuchungen konnten diese Thesen nicht uneingeschränkt bestätigt werden. Es ergaben sich sowohl theoretisch begründete Zweifel als auch konträre empirische Ergebnisse:

- * Informationsnachfrage kann durch vorhandenes (Experten-) Wissen substituiert werden;⁴⁾
- * hohe Informationsnachfrage ist auch Ausdruck von Unsicherheit, Zielunklarheit⁵⁾ und mangelnder Kenntnis des Informationsangebotes;⁶⁾
- * selbst-induzierte, dysfunktional wirkende Informationsüberflutung kann nicht ausgeschlossen werden;⁷⁾

- * andere (Informations-) Aktivitäten müssen hinzutreten, um das effizienzfördernde Potential der Informationsnachfrage zu nutzen: Die angeforderten Informationen müssen zweckmäßig gespeichert und problemgerecht verarbeitet werden, die verschiedenen Informationsaktivitäten sind aufeinander abzustimmen.⁸⁾
- * Die Effizienzwirkung ist nicht nur eine Frage der Menge, sondern auch der Qualität der Nachfrage und der beschafften Informationen.⁹⁾

Diese Differenzierungen schmälern nicht das Verdienst Wittes. Seine Arbeit zeichnet sich nicht nur durch eine innovative, reliable und valide Erhebungsmethode zur Messung der Informationsnachfrage im empirischen Feld aus. Witte und sein Team entwickeln auch eine neuartige Methode zur prozeßbegleitenden Messung des Informationsverhaltens unter experimentell kontrollierten Bedingungen, und zwar deutlich bevor sich solche Methoden auf breiterer Basis in der psychologischen Forschung über komplexe Probleme, der medizinischen verhaltensorientierten Forschung, der Marketingforschung und der verhaltensorientierten Rechnungswesenforschung durchsetzen.¹⁰⁾ Vor allem aber weist seine Schrift einen so erheblichen Theorieüberschuß auf, daß es rückblickend gesehen gar nicht möglich war, all seine "Hypothesen-Elemente" im Rahmen eines einzigen Forschungsprojektes abschließend zu prüfen.¹¹⁾

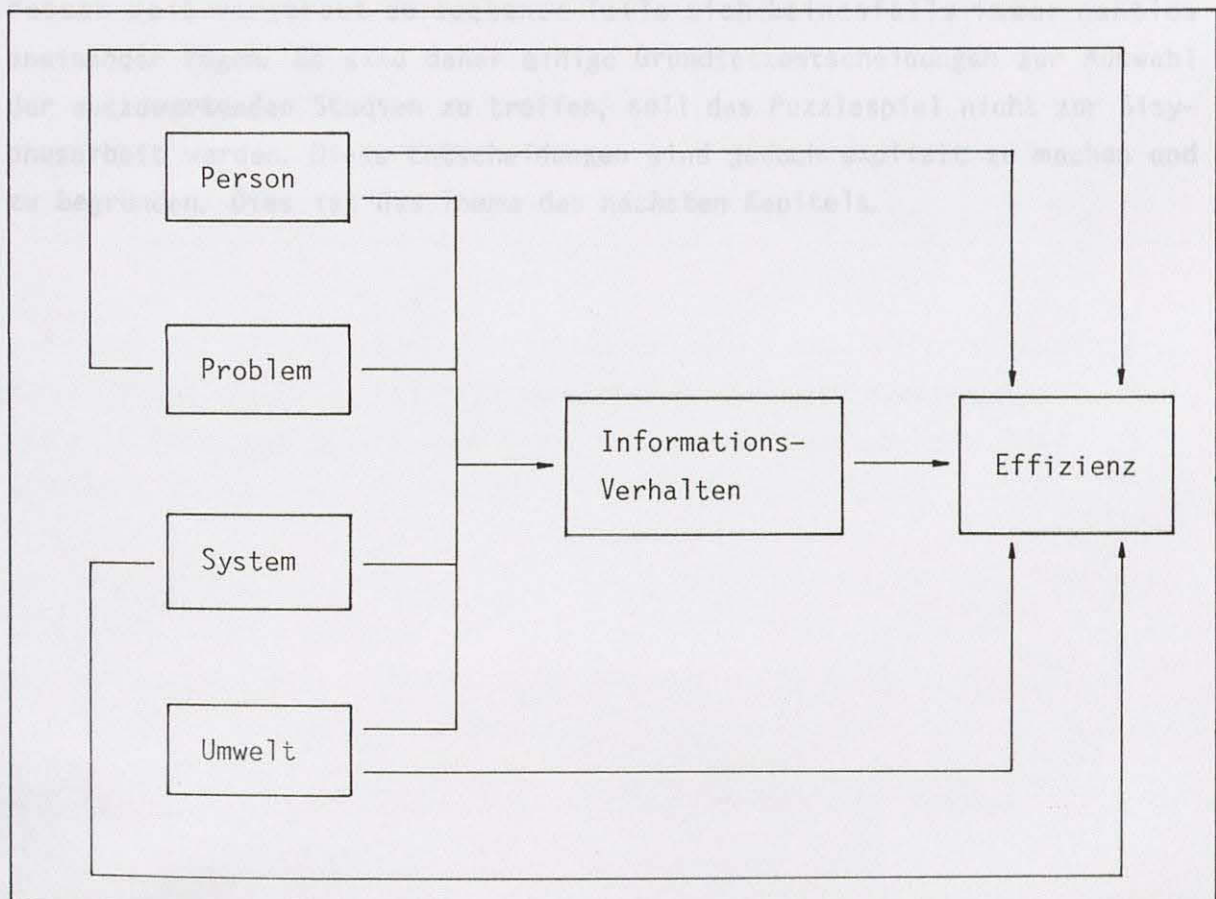
Es erscheint uns eine konsequente Fortführung der Arbeit Wittes, den von ihm vorgezeichneten Bezugsrahmen zu ergänzen und zu vertiefen und eine systematische Bestandsaufnahme der mittlerweile publizierten Arbeiten vorzunehmen. Dieses Unternehmen wird uns nicht nur in die amerikanische betriebswirtschaftliche Forschung, sondern auch in die neuere deutsche psychologische Problemlöseforschung führen, denn die Resonanz der deutschsprachigen betriebswirtschaftlichen Forschung auf die Arbeit von Witte ist vergleichsweise schwach. Mit Ausnahme der Arbeiten von Bronner¹²⁾, Brockhoff¹³⁾ und Hauschildt¹⁴⁾ hat es unseres Wissens keine weitere deutschsprachige empirische Studie gegeben, in der der Zusammenhang zwischen dem prozeßbegleitend gemessenen Informationsverhalten und der Effizienz eines komplexen betriebswirtschaftlichen Informationsproblems geprüft wurde.

Unserem erweiterten Bezugsrahmen liegt die Kontingenzhypothese zugrunde, daß die Wirkung der Informationsaktivität auf die Effizienz nicht generell behauptet werden sollte, sondern

- * die unterschiedlichen Anforderungen verschiedenartiger **Probleme**,
- * die unterschiedliche Leistungsfähigkeit und -bereitschaft verschiedenartiger **Personen**,
- * die unterschiedlichen Informations- und Entscheidungshilfen verschiedenartiger **Informationssysteme** und
- * den unterschiedlichen Kontexteinfluß verschiedenartiger **Umwelten**

berücksichtigen sollten. Die nachfolgende Abbildung veranschaulicht unser

Abbildung 2.1.: Bezugsrahmen zur Untersuchung der Zusammenhänge von Rahmenvariablen, Informationsverhalten und Effizienz



Grundkonzept. Durchaus vorstellbare Interaktionswirkungen der Merkmalsblöcke sind der Übersichtlichkeit halber weggelassen worden.

Diese Kontingenzhypothese ist nicht neu. Sie wird vielmehr in zahlreichen Bezugsrahmen, die zukünftige empirische Forschung lenken oder die Ergebnisse vergangener Forschung bewerten wollen, vertreten. Die nachfolgende Aufstellung stärker beachteter Bezugsrahmen, insbesondere aus der verhaltensorientierten Rechnungswesenforschung und der empirischen Forschung über computergestützte Informationssysteme, illustriert allerdings, daß die Einflußgrößen unterschiedlich stark differenziert werden.¹⁵⁾

Höchste Priorität hätte für uns eine Studie, die all diese Faktoren simultan berücksichtigt und womöglich noch ihre Wechselwirkungen untersucht. Eine solche "Super"-Studie gibt es natürlich nicht und kann es vermutlich auch gar nicht geben. Angesichts begrenzter Forschungskapazitäten muß man sich vielmehr auf bestimmte Realitätsausschnitte beschränken, die nach bestimmten Kriterien, mit bestimmten Verfahren und mit einem bestimmten Auflösungsgrad untersucht werden. Die Bestandsaufnahme und Integration bisher vorliegender Forschungsergebnisse wird somit zu einem mühsamen Puzzlespiel, dessen weit verstreut zu suchende Teile sich keinesfalls immer nahtlos aneinander fügen. Es sind daher einige Grundsatzentscheidungen zur Auswahl der auszuwertenden Studien zu treffen, soll das Puzzlespiel nicht zur Sisyphusarbeit werden. Diese Entscheidungen sind jedoch explizit zu machen und zu begründen. Dies ist das Thema des nächsten Kapitels.

Abbildung 2.2.: Relative Intensität der Beachtung der Faktoren Person, Problem, Informationssystem und Umwelt in theoretischen Bezugsrahmen

Theoretische Bezugsrahmen	Einflußfaktoren			
	Person	Problem	System	Umwelt
Mason/Mitroff (Management 1973)	XX	XX	XX	X
Mock (Information Structure 1973)	X	-	XX	X
Lucas (Information Systems 1975)	XX	X	XXX	XX
Ashton et al. (AAA-Report 1977)	XX	X	XX	-
Brockhoff (Beobachtungen 1977)	XX	-	XX	X
Dickson et al. (Minnesota Experim. 1977)	XX	X	XXX	X
Libby/Lewis (State of the Art 1977, 1982)	XX	X	XXX	X
Möllhoff (Informationsnachfrage 1978)	XXX	XX	XXX	XX
Brockhoff (Delphi-Prognosen 1979)	XX	X	XX	X
Zmud (Individual Differences 1979)	XXX	-	X	-
Ives/Hamilton/Davis (MIS 1980)	X	X	XX	XX
Ghani/Lusk (Information Processing 1982)	X	X	X	X
Courtney/DeSanctis/Kasper (MIS/DSS 1983)	X	XX	XXX	X
Hauschildt et al. (Geschäftsführung 1983)	X	XXX	XX	X
Mock/Vasarhelyi (Cognitive Style 1984)	XX	X	XX	-
Hauschildt (Graphische Unterstützung 1985)	X	X	X	X

Bedeutung der Eintragungen: Einflußfaktor wird nicht explizit genannt (-), exemplarisch angeführt (X), mit ausgewählten Befunden oder Hypothesen dargestellt (XX), mit zahlreichen Hypothesen und Befunden belegt (XXX)

2.1.2. Abgrenzung des Forschungsfeldes

2.1.2.1. Die Kriterien

Unsere Bestandsaufnahme konzentriert sich auf empirische Studien, die

- a) ein **komplexes Informationsproblem** untersuchen,
- b) (mindestens) eine betriebswirtschaftlich beachtliche **Effizienzgröße** erklären,
- c) das **Informationsverhalten** entweder als direkt (prozeßbegleitend) gemessene oder als unbeobachtete, intervenierende Variable zur Erklärung heranziehen,
- d) bis Redaktionsschluß **zugänglich** waren.
- e) Eine Einschränkung auf Studien, die bestimmte, besonders anspruchsvolle **methodische Standards** erfüllen, erfolgt nicht.

2.1.2.2. Komplexes Informationsproblem

Eine Beschränkung auf komplexe Informationsprobleme wird vorgenommen, um die zahlreichen Studien, die sich mit sehr einfachen, hoch abstrakten Aufgaben befassen, auszuschließen, denn ihre Ergebnisse können nur mit erheblichen Einschränkungen auf die betriebswirtschaftlichen Informationsprobleme, die uns in dieser Arbeit beschäftigen, übertragen werden. Unser Ausschluß betrifft auch zahlreiche Arbeiten aus der verhaltensorientierten amerikanischen Rechnungswesenforschung, in denen untersucht wird, wie eine vergleichsweise geringe Zahl vorgegebener Informationen (typischerweise 3 bis 7) zu einem klassifikatorischen Gesamturteil oder einer subjektiven Wahrscheinlichkeitsaussage verdichtet wird.¹⁶⁾ Trotz dieser Begrenzung unseres Forschungsfeldes verbleibt eine große Vielfalt unterschiedlicher Informationsprobleme. Die festgestellten Einflüsse auf die Effizienz sind daher nach den Kontextmerkmalen Art und Komplexität des Informationsverhal-

tens zu differenzieren. Anzumerken ist, daß die meisten gefundenen Studien sich auf **individuelle** Informationsprozesse beziehen. Eine a priori Beschränkung auf solche Studien wollten wir jedoch nicht vornehmen, denn dies hätte zur Folge gehabt, daß wir die für unsere Studie wichtigen Arbeiten von Witte, Brockhoff und Hauschildt "hinausdefiniert" hätten. Wir werden bei der Würdigung der Ergebnisse auf diese Besonderheit eingehen.

2.1.2.3. Effizienz als abhängige Variable

Unsere Arbeit orientiert sich am Gestaltungsziel der Betriebswirtschaftslehre. Deshalb konzentrieren wir uns auf Arbeiten, in denen die **Wirkungen** des Informationsverhaltens untersucht werden. Abzugrenzen ist, welche Wirkungen dies sein sollen. Unstrittig ist, daß alle Arbeiten, in denen ökonomische, monetäre Effizienzgrößen untersucht werden, zu unserer Grundgesamtheit gehören. Darüber hinaus halten wir es für zweckmäßig, Studien zu berücksichtigen, die Effizienzmaße verwenden, welche in einer positiven Beziehung zu ökonomischen Effizienzgrößen stehen. Dies sind Maßgrößen der Prognosegüte, der Urteilsqualität und der Entscheidungsqualität. Sie werden vornehmlich dann verwendet, wenn die Forscher sich auf einen Teilprozeß einer umfassender definierten Entscheidung konzentrieren, monetäre Effizienzmaße nicht beschaffen können, oder wenn sie wegen einer zu großen Zurechnungsproblematik der komplexen (!) Entscheidung auf monetäre Effizienzmaße bewußt verzichten.

Ausgeschlossen werden alle Studien, in denen ausschließlich Zufriedenheitsmaße, Maße der subjektiven Urteilsconfidenz oder Maße des Zeitverbrauches als abhängige Größen verwendet werden. Diese Größen sind für sich genommen nicht beachtlich für unsere Auswertungszwecke.¹⁷⁾ Als Prädiktoren ökonomischer Effizienzgrößen sind sie zu unzuverlässig.

Anzumerken ist, daß es zahlreiche Studien gibt, die das Informationsverhalten, namentlich die Menge an beschafften Informationen, als eine "Effizienzgröße" ansehen. Dieser Auffassung folgen wir nicht. Vom Informationsverhalten können sowohl positive als auch negative Einflüsse auf uns interessierende Effizienzgrößen ausgehen. Sie differenziert zu erforschen, ist

gerade das Ziel unserer Studie.

Ein wichtiges Ergebnis unserer Bestandsaufnahme dürfen wir vorwegnehmen: In der großen Mehrheit der referierten Studien werden monetäre Effizienzgrößen verwendet. Unsere Abgrenzungsentscheidung, auch nicht-monetär bestimmte ökonomische Effizienzmaße zu verwenden, führt faktisch gesehen nicht zu einer "Aufweichung" des betriebswirtschaftlichen Problemgehaltes, wie sie von manchen Autoren des Faches befürchtet wird.¹⁸⁾

2.1.2.4. Informationsverhalten als unabhängige Variable

Bei der Analyse des Zusammenhanges zwischen Informationsverhalten und Effizienz haben sich drei Paradigmen in der Forschungspraxis durchgesetzt:

1. Im Paradigma I wird ein prozeßbegleitend **gemessenes** Informationsverhalten zur Effizienz in Beziehung gesetzt. Dieses Paradigma wurde in der deutschen betriebswirtschaftlichen Forschung von Witte und seinen akademischen Schülern und von Brockhoff angewandt.¹⁹⁾ Aus der amerikanischen Forschung ist die Arbeit von Shields hervorzuheben.²⁰⁾ In jüngerer Zeit wird das Paradigma in Arbeiten der deutschsprachigen psychologischen Problemlöseforschung angewandt, die Verlauf und Erfolg von Denkprozessen bei computer-gestützten Simulationen komplexer Managementprobleme erforschen.²¹⁾ Ein viertes Anwendungsfeld stellen medizinische Diagnoseprozesse dar, die zahlreiche Parallelen zu betriebswirtschaftlichen Beurteilungsprozessen aufweisen. Umfangreiche Forschungsarbeiten hierzu wurden vor allem von Elstein und seinen akademischen Kollegen durchgeführt.²²⁾ Anzumerken ist, daß sich alle vier Forschungsrichtungen unabhängig voneinander entwickelten und bis heute praktisch keinen Bezug aufeinander nehmen.²³⁾

2. Im Paradigma II werden Merkmale des **Informationssystems** systematisch variiert. Man geht davon aus, daß Gestaltungsparameter wie Menge, Aggregationsgrad und Präsentationsform der angebotenen Daten Informationsverhalten und Effizienz beeinflussen. Experimentelle Studien dieser Art werden sowohl in der verhaltensorientierten Rechnungswesenforschung ("behavioral accounting")²⁴⁾ als auch in der Forschung über Management-Informations- und Ent-

scheidungs-Unterstützungs-Systeme ("MIS/DSS")²⁵⁾ durchgeführt. Bekannt geworden sind die "Minnesota Studies".²⁶⁾ Hervorzuheben sind auch die Arbeiten von Lucas²⁷⁾ und Benbasat.²⁸⁾ Bei beiden Forschungsgebieten finden sich fast ausschließlich englischsprachige Veröffentlichungen. Die deutsche Forschung weist ein erhebliches, beständig wachsendes Defizit auf.

3. Im Paradigma III werden Eigenschaften der **Person** zum Informationsverhalten und zur Effizienz in Beziehung gesetzt. Besonders häufig werden sog. "Kognitive Stile"²⁹⁾ und die sog. "Test"-Intelligenz³⁰⁾ als Prädiktoren verwendet. Die erste Variablengruppe wird vor allem in der amerikanischen betriebswirtschaftlichen Forschung untersucht. Untersuchungen zur zweiten Gruppe werden vornehmlich von deutschsprachigen Psychologen vorgelegt. Anzumerken ist, daß es eine Reihe empirischer Studien gibt, in denen man die **Wechselwirkungen** von Persönlichkeits- und System- oder Aufgabenmerkmalen auf die Effizienz untersucht. Hinter den Arbeiten der amerikanischen Managementforschung steht dabei die Idee, Informationssysteme zu entwickeln, die auf Persönlichkeitstypen mit einem bestimmten kognitiven Stil zugeschnitten sind. Die psychologische Problemlöseforschung möchte herausfinden, welche Anforderungen komplexe Probleme stellen und wieweit diese den mit Intelligenztests gemessenen Eigenschaften entsprechen.³¹⁾

Für unsere Arbeit haben Studien, die Paradigma I folgen, eine besondere Relevanz. Es erscheint uns jedoch nicht zweckmäßig, unsere Bestandsaufnahme auf diese Untersuchungen zu beschränken.

Merkmale des Informationssystems wie z. B. die Menge an angebotenen Informationen, oder Eigenschaften einer Person, wie z. B. ihre "Kognitive Komplexität", besitzen eine empirisch nachgewiesene Beziehung zum Informationsverhalten. So zeigen verschiedene Studien aus der Marketingforschung und der psychologischen Entscheidungsforschung, daß die absolute Menge an beschafften Informationen mit der Angebotsmenge steigt und die an der Angebotsmenge relativierte Nachfrage sinkt.³²⁾ Nach der vielzitierten Untersuchung von Schroder, Driver und Streufert entfalten kognitiv "komplexe" Individuen ein höheres Niveau der Informationsverarbeitung als "einfache".³³⁾ Ihre Befunde wurden in zahlreichen Nachfolgestudien bestätigt und auf andere Dimensionen des Informationsverhaltens ausgedehnt.³⁴⁾ Es bietet

sich daher an, empirische Studien zum Paradigma II und III auszuwerten, weil sie Hinweise auf die **Ursachen** unterschiedlichen Nachfrageverhaltens liefern. Sie erfüllen damit eine wichtige Aufgabe für die **Erklärungsaufgabe** der Betriebswirtschaftslehre, denn bei einer reinen Analyse des Zusammenhangs von Informationsverhalten und Effizienz bleibt die Frage offen, ob beobachtete Verhaltensunterschiede tatsächlich als "Ursache" für festgestellte Effizienzunterschiede verantwortlich gemacht werden können. So kann man beispielsweise bei einem gleichzeitigen Auftreten eines hohen Niveaus der Informationsverarbeitung und einer hohen Entscheidungseffizienz fragen, ob dieser Zusammenhang auf einer höheren kognitiven Komplexität der erfolgreicheren Entscheidungsträger beruht.

Doch die Aussicht, eine Verlängerung der Kausalkette zu erzielen, ist nicht das ausschlaggebende Motiv für eine kritische Bestandsaufnahme der Studien, die Paradigma II und III anwenden. Selbst wenn dies der Fall wäre, müßte berücksichtigt werden, daß die bisher berichteten Einflüsse von System- und Persönlichkeitsmerkmalen auf das Informationsverhalten insgesamt gesehen noch zu wenig belegt, zu uneinheitlich und zu schwach ausgeprägt sind.³⁵⁾ Der zweite Grund für eine Bestandsaufnahme ist vielmehr die von uns stärker gewichtete **Gestaltungsaufgabe** der Betriebswirtschaftslehre:

- * Die untersuchten **Merkmale des Informationssystems** sind **Gestaltungsparameter**, die man direkt verändern kann. Mit Hilfe geeigneter faktorieller Experimente kann man die angestrebten Wirkungen vergleichsweise leicht überprüfen. Die Ergebnisse solcher Experimente lassen sich kausal interpretieren.
- * **Psychometrische Tests zur Messung von Persönlichkeitsvariablen** sind empirisch geprüfte, vergleichsweise leicht anwendbare Meßinstrumente, auf denen man eine Personalrekrutierungs- und -schulungspolitik aufbauen kann, die allen Beteiligten nützt - vorausgesetzt, diese Instrumente erfüllen tatsächlich die für den Anwendungszweck erforderliche Meßleistung.

- * Die **prozeßbegleitende Messung des Informationsverhaltens** ist demgegenüber eine sehr aufwendige Methode. Die bisher durchgeführten Analysen werfen nicht zu unterschätzende Probleme bezüglich der kausalen Interpretation der gefundenen Effekte auf. Man kann das Informationsverhalten nur **indirekt** über **andere** Instrumente beeinflussen. Welche Instrumente dies sind und wie sie wirken, ist bisher noch nicht systematisch erforscht worden. Deshalb ist das Kosten-Nutzen-Verhältnis von Wirkungsaussagen über das Informationsverhalten auf die Effizienz sehr viel ungünstiger als von Wirkungsaussagen, die sich auf System- oder Personenmerkmale beziehen.

Prozeßbegleitende Meßverfahren sind nach unserem Eindruck – mit Blick auf eine stark hervorgehobene Erklärungsaufgabe – z. T. mit erheblichen Vor-schußlorbeeren bedacht worden.³⁶⁾ Betrachtet man die bisherige Praxis der **Anwendung** solcher Verfahren in der verhaltensorientierten Rechnungswesenforschung, dann stellt man fest, daß die meisten bisher publizierten Studien explorative Arbeiten sind, die sich in einer zwar sehr detaillierten, aber doch rein deskriptiven Darstellung erschöpfen.³⁷⁾ Hinzu kommt, daß man sich wegen der aufwendigen Meßverfahren meistens auf kleine Stichproben, die deutlich unter 30 Versuchspersonen liegen, beschränkt. Nur in wenigen Studien werden quantitative Meßwerte gebildet, um Existenz- und Kontingenzhypothesen über den Prozeßverlauf zu überprüfen.³⁸⁾ Um die noch selteneren Studien zu finden, in denen **Effizienzhypothesen** zum Einfluß eines prozeßbegleitend gemessenen Informationsverhaltens abgeleitet und geprüft werden, muß man den Suchraum erheblich ausweiten und auch Ergebnisse aus Nachbardisziplinen hinzunehmen.³⁹⁾

Man muß sich allerdings kritisch fragen, ob es sich überhaupt lohnt, solch aufwendige Prozeßstudien durchzuführen, wenn das Ziel eine Verbesserung der Effizienzprognose und -erklärung ist. Bevor man die bisherigen Methoden bei größeren Stichproben anwendet, sollte zunächst einmal prüfen, ob die bereits vorliegenden Studien tatsächlich **besser** geeignet sind, Effizienzunterschiede zu erklären, als solche Arbeiten, die ausschließlich System- und Personenmerkmale zur Effizienzprognose heranziehen. Dies ist unser zweiter Grund, eine systematische Bestandsaufnahme der Studien zum Paradigma II und III durchzuführen.

Allerdings wollen wir die Frage nicht so undifferenziert stellen: Man will doch wissen, **welche** System- und Personenvariablen die Effizienz beeinflussen, **warum** sie die Effizienz verringern oder erhöhen und welche Rolle das **Informationsverhalten** dabei spielt. Deshalb bedarf es einer differenzierten Bestandsaufnahme zahlreicher Einflußfaktoren und ihrer Wechselwirkungen. Unser dritter Grund für eine Bestandsaufnahme ist daher das Bestreben, die bisher ziemlich isoliert durchgeführten Untersuchungen einer leistungsfähigeren, **integrativen Theorie** zuzuführen.

Anzumerken ist, daß es auch eine Reihe empirischer Studien gibt, in denen Merkmale des **Problems** zu Informationsverhalten und Effizienz in Beziehung gesetzt werden. Wir werden diese Studien nicht referieren, da sie sich meistens auf das Informationsverhalten als abhängige Variable beschränken. Dies läßt sich damit erklären, daß es sehr schwierig, wenn nicht gar unmöglich ist, ein Effizienzmaß zu finden, das sich für sehr unterschiedliche Aufgaben gleich gut eignet. Insgesamt gesehen beschränken sich die uns bekannten Studien auf jeweils ausgewählte Typen von Informationsproblemen. Innerhalb eines bestimmten Problemtyps werden nur wenige, ausgewählte Parameter variiert. Von einer geschlossenen Theorie des Einflusses von Problem-Merkmalen ist man noch weit entfernt.⁴⁰⁾

2.1.2.5. Verfügbarkeit der Studien

Eine systematische Bestandsaufnahme bedeutet nicht, daß man alle irgendwann und irgendwo publizierten Studien berücksichtigt. Die methodischen Schriften zur Meta-Analyse⁴¹⁾ empfehlen vielmehr, sich auf bestimmte Forschungsfragen zu konzentrieren und **hypothesengeleitet** vorzugehen.⁴²⁾ Wir folgen dieser Empfehlung und konzentrieren uns auf solche Personen-, System- und Verhaltensmerkmale, die nach Meinung der Literatur einen beachtlichen Einfluß auf die Effizienz ausüben. Was "beachtlich" ist, kann man sowohl an der Qualität der Hypothesenableitungen als auch an der Qualität und Quantität der empirischen Prüfungen ablesen. Nur bezüglich dieser relevanten Teilmenge von empirischen Studien stellt sich die Frage nach der "Verfügbarkeit".

Als **"verfügbar"** sehen wir alle in deutscher oder englischer Sprache publizierten Studien an, die in Kieler Universitätsbibliotheken oder im innerdeutschen Fernleihverkehr bezogen werden können. **Dissertationen** werden nur dann berücksichtigt, wenn sie über die genannten Stellen beschafft werden konnten. Arbeitspapiere, Manuskripte und ähnliche sogenannte **"graue" Literatur** wird nur dann verwendet, wenn die Autoren keine Einwendungen gegen eine Zitation erheben, die Schriften im öffentlichen innerdeutschen Leihverkehr bezogen werden können oder in das persönliche Eigentum des Verfassers übergegangen sind, so daß er sie interessierten Dritten auf Wunsch gegen geeignete Konditionen zur Verfügung stellen kann.

Unsere restriktive Behandlung der sog. "grauen" Literatur steht in gewissem Widerspruch zu manchen Empfehlungen der Literatur zur Meta-Analyse. So schlagen beispielsweise Fricke und Treinies vor, solche Quellen nicht a priori auszuschließen, weil man sie für methodisch weniger qualifiziert hält.⁴³⁾ Die Autoren weisen ferner daraufhin, daß bis zur Veröffentlichung in Zeitschriften und Sammelwerken eine erhebliche Zeitspanne verstreicht. Man kann hinzufügen, daß gelegentlich überhaupt keine Publikation erfolgt und daß die Arbeitspapiere z. T. wertvoller sind als Zeitschriftenaufsätze, weil sie wichtige zusätzliche Informationen enthalten, die eine differenzierte Beurteilung unterstützen. Wir haben jedoch nicht wegen eines grundsätzlichen Mißtrauens gegenüber einer möglicherweise schlechteren Qualität so wenige Arbeitspapiere verarbeitet. Maßgeblich war vielmehr die Tatsache, daß sich eine Beschaffung der meist in den Vereinigten Staaten von Amerika verbreiteten Arbeitspapiere sehr schwierig gestaltete.⁴⁴⁾ Wir wollen aber nur solche Forschungsergebnisse berücksichtigen, die wir selbst prüfen und interessierten Dritten zugänglich machen können, damit diese auch tatsächlich auf unserer Bestandsaufnahme aufbauen können.

Zur Nachvollziehbarkeit der Bestandsaufnahme ist kurz darzustellen, wie die **Literatursuche** erfolgte.⁴⁵⁾

1. Unsere Literatursuche erstreckte sich in erster Linie auf die namhaften deutsch- und englischsprachigen **betriebswirtschaftlichen Zeitschriften**, in denen häufiger uns interessierende Arbeiten publiziert werden. Besonders

ergiebig waren "Management Science", "Decision Science", "Accounting, Organizations and Society", "Journal of Accounting Research", "The Accounting Review" und deren Supplemente. Weitere Beiträge fanden wir in "IEEE Transactions on Systems, Man and Cybernetics" und in "Management Information Systems Quarterly". In deutschsprachigen Zeitschriften scheinen die hier untersuchten Zusammenhänge offensichtlich "kein Thema" zu sein, denn aus ihnen konnten wir nur sehr wenige Beiträge entnehmen. Abzuwarten bleibt, welche Beiträge die neugegründeten Zeitschriften "Computers in Human Behavior" und "Human Computer Interaction" für das hier betrachtete Forschungsfeld anbieten.

2. Die **psychologischen Zeitschriften** konnten wir aus Kapazitätsgründen nicht vollständig und präzise überprüfen. Man darf jedoch angesichts der bisherigen stürmischen Entwicklung davon ausgehen, daß in den Zeitschriften "Diagnostica", "Sprache und Kognition", "Organizational Behavior and Human Decision Processes" (früher "Performance" statt "Decision Processes"), "Psychologische Rundschau", "Zeitschrift für Psychologie" und "Zeitschrift für experimentelle und angewandte Psychologie" auch in Zukunft interessante Arbeiten zum Zusammenhang von Person, System, Informationsverhalten und Effizienz bei komplexen Managementproblemen publiziert werden.

3. **Dissertationen** wurden überprüft an Hand des "Dissertation Abstracts" wobei wir uns auf die Rubrik "Accounting" beschränkten. Die Beschaffungsbemühungen endeten leider häufig mit dem Hinweis "Bestellung erfolglos geblieben". Neuere deutschsprachige Dissertationen und Habilitationsschriften wurden an Hand des "Hochschulschriftenverzeichnisses" der Deutschen Bibliothek in Frankfurt am Main überprüft.

Weitere Hinweise entnahmen wir den **Literaturverzeichnissen** der im Text zitierten Monographien, Sammelwerke und Zeitschriftenaufsätze.

Eine On-Line-Recherche haben wir nicht vorgenommen, da unsere Kosten-Nutzen-Erwartung zu ungünstig ausgeprägt war.

2.1.2.6. Keine Vorauswahl aufgrund methodischer Standards

Es dürfte unstrittig sein, daß man die methodische Qualität der empirischen Studien bei einer Bestandsaufnahme berücksichtigt. Dies bedeutet jedoch nicht, daß man sich von **vornherein** auf Arbeiten beschränkt, die als methodisch besonders anspruchsvoll angesehen werden. Eine solche Strategie könnte insbesondere folgende negativen Auswirkungen haben:

1. Untersuchungen, die vergleichsweise einfache, statische Informationsprobleme analysieren, könnten in unserer Bestandsaufnahme besonders stark vertreten sein, weil sich solche Probleme mit einer größeren internen Validität des Designs erforschen lassen. Die für uns besonders interessanten Arbeiten, die vergleichsweise komplexe Probleme bearbeiten, blieben auf der Strecke.
2. Untersuchungen, die methodische Mängel aufweisen, aber relativ starke Effekte aufdecken, würden vernachlässigt. Dies wäre vor allem dann problematisch, wenn sich die starken Effekte nicht gerade wegen, sondern trotz der vermuteten Mängel zeigen.
3. Die Beurteilung der methodischen Qualität einer Studie ist ein komplexer Prozeß, der sich auf viele Kriterien stützt. Wie diese Kriterien zu gewichten sind, und welche Bedeutung sie im konkreten Einzelfall haben, ist häufig eine Frage der subjektiven Einschätzung. Dabei gibt es häufig einen ganz erheblichen Ermessensspielraum, weil klare Hinweise auf das Ausmaß eines bestimmten methodischen Problems fehlen und weil es noch keine empirisch abgesicherte "Artefakt"-Theorie gibt, aus der man zuverlässige Schätzungen ableiten kann, wie **stark** die Auswirkungen solcher Mängel sind.

Wir folgen deshalb den Empfehlungen der Schriften zur Meta-Analyse, keine strenge Vorauswahl aufgrund methodischer Kriterien vorzunehmen, sondern stattdessen möglichst viele Studien zu berücksichtigen und diese anschließend, in einer für den Leser **nachvollziehbaren** Art und Weise unterschiedlich zu gewichten. Nur so kann das Potential einer Meta-Analyse genutzt werden: Es gibt wohl keine empirische Studie, die nicht irgendeinen

Mangel aufweist, den man als beachtlich für eine bestimmte Fragestellung ansehen kann. Andererseits ist es sehr unwahrscheinlich, daß alle Studien zu einem bestimmten Problem genau dieselben Mängel aufweisen. Wenn ein Effekt besteht, dann sollte er sich deshalb im Durchschnitt aller vergleichbaren Studien zeigen. Wenn spezifische Mängel das Auftreten dieses Effektes besonders stark beeinträchtigen, dann sollte bei allen Studien, die diesen methodischen Mangel aufweisen, ein geringerer Effekt zu beobachten sein. Für uns sind methodische Einwände **Alternativhypothesen**. Man kann und sollte sie genauso prüfen wie andere Hypothesen auch: durch Analyse ihrer logischen Konsistenz und durch Konfrontation mit geeigneten empirischen Daten. Um genau dies zu ermöglichen, verzichten wir auf eine Vorselektion aufgrund bestimmter methodischer Kriterien.

2.1.3. Beurteilung der empirischen Studien

Es gibt unterschiedlich anspruchsvolle Möglichkeiten, empirische Forschungsergebnisse zu berücksichtigen:⁴⁶⁾

Niveau 1: Man bezieht sich auf ausgewählte, häufig zitierte Studien.

Niveau 2: Man bezieht sich auf eine ganze Reihe empirischer Vorarbeiten, weist auf deren Inkonsistenzen und Defizite hin und präsentiert dann eine eigene Studie, die diese Defizite überwinden soll.

Niveau 3: Man führt eine systematische Bestandsaufnahme durch, die die bisher durchgeführten Arbeiten nach Forschungsfragen ordnet und ihre Ergebnisse nach verschiedenen Kriterien systematisch auswertet. Für diese Art der Bestandsaufnahme setzt sich in neuerer Zeit die Bezeichnung "Meta-Analyse"⁴⁷⁾ durch. Sie zeichnet sich dadurch aus, daß man die Bestandsaufnahme als eine eigenständige empirische Arbeit begreift und spezielle Methoden zur Gewinnung und Verarbeitung der Einzelergebnisse einsetzt. Da wir nur eine Teilmenge der exponentiell wachsenden Menge an vorgeschlagenen Techniken nutzen, weil die hier ausgewerteten Studien sehr heterogen sind und häufig unvollständige Angaben machen, wollen wir lediglich von einer "systematischen Bestandsaufnahme" sprechen.⁴⁸⁾

In der bisherigen Forschungspraxis werden meistens nur die beiden ersten Anspruchsniveaus realisiert.⁴⁹⁾ Ein Verharren auf dem ersten Niveau kann dazu führen, daß Hypothesen, die schon lange nicht mehr in der ursprünglichen Form aufrechterhalten werden können, dennoch weiter verbreitet und verfestigt werden.⁵⁰⁾ Mit dem Erreichen der zweiten Stufe wird zwar eine differenziertere Betrachtung vorgenommen. Man beschränkt sich aber häufig darauf, lediglich auf Inkonsistenzen hinzuweisen und forscht nicht systematisch nach möglichen Ursachen für divergierende Ergebnisse.

Deshalb ist eine systematische Bestandsaufnahme anzustreben. Sie zeichnet sich dadurch aus, daß man

- * auch nicht "signifikante" Ergebnisse berücksichtigt,
- * eine Beurteilung nach einheitlichen Kriterien vornimmt und
- * mögliche Erklärungen für divergierende Ergebnisse empirisch prüft – soweit es die Angaben der referierten Studien zulassen.

Wir streben eine systematische, nachvollziehbare Bestandsaufnahme an und müssen daher erläutern, nach welchen Kriterien die Beurteilung der Studien vorgenommen wird.

1. Zur Beurteilung der empirischen Evidenz, die zu einer bestimmten Forschungshypothese vorgelegt wird, nehmen wir zunächst eine **dreiklassige Abstufung** nach **Verlauf** und **Signifikanz** der Ergebnisse vor:

"f.": **falsifizierende** Befunde widersprechen der Hypothese durch konträre Verläufe oder fehlende Zusammenhänge,

"(b.)": eine **eingeschränkte Bestätigung** liefern Befunde mit einem hypothesenkonformen, aber nur tendenziell signifikanten Verlauf (0,05 kl. p kl. 0,10) oder mit deutlich hypothesenkonformem Verlauf ohne Signifikanzangabe.

"b.": Eine **volle Bestätigung** geben Befunde mit hypothesenkonformem Verlauf und signifikanten Unterschieden.

2. Als ergänzende Informationen werden die ermittelten **Effektstärken** und der **Stichprobenumfang** der Studien herangezogen.

3. Die Beurteilung stützt sich ferner auf die **Reliabilität** und **Validität** der Operationalisierung der unabhängigen und abhängigen Variablen.

4. Eine Detailanalyse der einzelnen Studien wird nur dann vorgenommen, wenn begründeter Anlaß besteht, daß Designspezifika für divergierende Ergebnisse verantwortlich gemacht werden können. Von dieser Möglichkeit wird sparsam Gebrauch gemacht: Es ist nicht unsere Absicht, alle beobachteten Divergenzen durch Ex-Post-Faktum-"Theorien" zu "erklären".

Zu diesen Beurteilungskriterien sind einige Anmerkungen erforderlich.

Die Verwendung von Signifikanztests⁵¹⁾ ist nicht für einzelne empirische Studien, sondern auch für Meta-Analysen⁵²⁾ kritisch beurteilt worden. Unsere Verwendung von Signifikanztests als Beurteilungskriterium beruht vor allem darauf, daß es sich um ein **häufig publiziertes Mindestkriterium** handelt. Würden wir jene Studien, die lediglich berichten, daß keine signifikanten Unterschiede oder Zusammenhänge festgestellt wurden, ausschließen, so würde dies zu einer systematischen Überschätzung des Gesamtergebnisses führen. Man muß jedoch einen besonderen Defekt der sog. "Auszählmethode" hervorheben: Die Wahrscheinlichkeit für einen sog. "Fehler zweiter Art", der darin besteht, irrtümlich von einem fehlenden Zusammenhang auszugehen, obwohl ein solcher besteht, kann bei der Auszählmethode sehr hoch ausfallen. So zeigen insbesondere Hedges und Olkin, daß bei kleinen Stichproben und kleinen Effekten in der Grundgesamtheit die Wahrscheinlichkeit für einen Fehler zweiter Art nicht etwa sinkt, sondern ziemlich rasch gegen 1 konvergiert.⁵³⁾

Hieraus ziehen wir folgenden Schluß:

- * Wenn die einfache Mehrheit der Studien ein signifikantes und in die gleiche Richtung weisendes Ergebnis besitzt, dann reicht bereits die Auszählmethode aus, eine Beurteilung des Gesamteffektes vorzunehmen.
- * Ergibt sich ein hoher Anteil nicht signifikanter Ergebnisse, dann ist zu prüfen, ob dies an einem zu geringen Stichprobenumfang liegt oder, ob es sich um einen allenfalls schwachen Zusammenhang in der Grundgesamtheit handelt.
- * Ergibt sich ein hoher Anteil von signifikanten, aber verlaufskonträren Ergebnissen, dann ist zu prüfen, ob es bestimmte Gemeinsamkeiten innerhalb der homogen ausgeprägten Gruppen von Studien gibt, die den Unterschied zwischen den Gruppen von Studien erklären können. Nach Hedges und Olkin ist es nämlich bei den üblichen Stichprobenumfängen extrem unwahrscheinlich, daß man bei einem fehlenden Zusammenhang in der Grundgesamtheit einen hohen Anteil signifikanter richtungskonträrer Mittelwertunterschiede entdeckt.⁵⁴⁾

Diese Verfeinerungen des Beurteilungsverfahrens sollen nicht darüber hinwegtäuschen, daß die Auszählmethode ein recht grobes Verfahren ist, weil sie sich nur auf eine publizierte Mindestinformation stützen kann. Wir werden deshalb nach den etwa bei Fricke und Treinies vorgeschlagenen Methoden versuchen, nachträglich Effektstärken zu schätzen. Dabei ist jedoch zu bedenken, daß es sich um teilweise recht grobe Abschätzungen handelt, die nur eine Größenordnung des "wahren" Effektes vermitteln können. Aus Gründen der Vergleichbarkeit werden wir die Effektstärken bei monotonen Zusammenhängen als Korrelationskoeffizienten und bei (umgekehrt) u-förmigen Verläufen als erklärte Varianzen ausdrücken.⁵⁵⁾ Nach Cohen, auf den sich viele Sozialwissenschaftler beziehen, sind Produktmomentkorrelationen von 0,10 als "klein", solche von 0,30 als "mittel" und solche von 0,50 als "groß" anzusehen.⁵⁶⁾ Da es bei unserer Bestandsaufnahme darum geht, Einflußfaktoren zu identifizieren, die für eine Effizienzprognose als beachtlich angesehen werden, sollte der Effekt schon eine "mittlere" Stärke im Sinne von Cohen aufweisen. Wenn wir einen Einflußfaktor als "unbeachtlich" klassifizieren, dann nehmen wir bei unserer Auszählmethode durchaus in Kauf, daß ein von Null verschiedener aber "kleiner" Effekt in der Grundgesamtheit

besteht. Aus diesem Grunde verzichten wir auch darauf, etwa die von Rosenthal diskutierten Methoden zur Schätzung eines gemeinsamen Signifikanzniveaus aus zahlreichen Studien zu bestimmen.⁵⁷⁾ Es kommt uns nicht darauf an, einen kleinen Effekt mit einer hohen Schlußsicherheit nachzuweisen. Wir wollen lediglich verhindern, daß wir einen in der Grundgesamtheit bestehenden mittleren bis großen Effekt irrtümlich als nicht existent bewerten.

Der Ausweis der Effektstärken soll einen Eindruck von der praktischen Bedeutsamkeit der Befunde für prognostische Zwecke geben. Wir sind uns der Tatsache bewußt, daß die Effektstärken von den Designspezifika der einzelnen Studien abhängen und eine induktive Verallgemeinerung nicht unproblematisch ist.⁵⁸⁾ Sie wird auch von uns nicht angestrebt. Der Leser soll vielmehr eine übersichtliche Gesamtinformation erhalten, um selbst beurteilen zu können, wie groß nach seiner Meinung ein mittlerer Effekt bei einer bestimmten (von ihm gebildeten) Art von Studien ist und wie groß die **empirischen** Abweichungen sind. Effektstärken sind vor allem ein nützliches Hilfsmittel zur Beurteilung eines Forschungsfeldes, in dem viele Studien mit kleinen Stichprobenumfängen durchgeführt werden. Wenn sich nämlich mehrheitlich schwache Effektstärken zeigen, dann kann man daraus schließen, daß die seltene Verwerfung der Nullhypothese nicht auf einer zu geringen Teststärke sondern auf einer zu geringen Effektstärke beruht. Damit wird überhaupt erst eine Falsifizierbarkeit (!) von Hypothesen möglich, denn aus der Annahme der statistischen Nullhypothese ist bei gängiger Testpraxis die Ablehnung der wissenschaftlichen Hypothese abzuleiten. Deshalb muß man wissen, worauf eine Nicht-Verwerfbarkeit der statistischen Nullhypothese beruht. Diese Ursachenklärung ist unseres Erachtens erst im Rahmen einer systematischen Bestandsaufnahme mehrerer Studien möglich.

Zur Nachvollziehbarkeit unserer Bewertung werden die wichtigsten Ergebnisse der einzelnen Studien tabellarisch dargestellt. Dabei werden die Aufgabenstellung, Zahl und Art der untersuchten Subjekte, die Operationalisierungen der unabhängigen und abhängigen Variablen und die Befunde in knapper Form zusammengefaßt. Unsere Bewertung stützt sich auf das jeweils umfassendste Effizienzmaß und die größte Auswertungsgesamtheit einer Studie. In den Untersuchungen ausgewiesene Befunddifferenzierungen nach Subpopulationen werden nur dann verwendet, wenn es sich um Aufspaltungen handelt, die für

unsere Auswertung beachtlich sind. Maßgeblich hierfür ist die Überlegung, eine möglichst große Teststärke zu erhalten.⁵⁹⁾ Der in der Spalte "Subjekt" angegebene Stichprobenumfang bezeichnet die effektive Anzahl an Untersuchungseinheiten, die dem jeweils referierten Befund zugrundeliegen.

In diesem Kapitel werden die Ergebnisse empirischer Studien zum Einsatz des Informationssystems auf die Effizienz komplexer betrieblicher Entscheidungsgewürdigt. Dabei werden zunächst die klassische Abgrenzungsparameter Menge und Aggregationsgrad der angebotenen Informationen dargestellt. Anschließend werden die Wirkungen der vornehmlich in jüngster Zeit gegen die Verteilten Medien und Präzentationsformate untersucht. Das Kapitel wird abgeschlossen mit Gestaltungsgrößen eines aktiven Informationssystems. Aufbau und formale Entscheidungsregeln.

Im Mittelpunkt unserer Bestandsaufnahme stehen Laborexperimente, in denen klassische Entscheidungsgrößen als abhängige Variable betrachtet werden, und zwar bei individueller Entscheidungsfindung. Diese Untersuchungen werden in der Regel als computergestützte Entscheidungsfindung durchgeführt. Die dabei ermittelten Ergebnisse methodischer Natur sind durch den Vergleich der gleichfalls ausgewerteten Ergebnisse der klassischen Entscheidungsfindung zu beurteilen. Die Ergebnisse der Experimente werden in der Regel in Form von Entscheidungsregeln dargestellt, die die Entscheidungsfindung unterstützen. Diese Regeln werden in der Regel in Form von Entscheidungsregeln dargestellt, die die Entscheidungsfindung unterstützen. Diese Regeln werden in der Regel in Form von Entscheidungsregeln dargestellt, die die Entscheidungsfindung unterstützen.

2.2. Analyse des Entscheidungsprozesses

2.2.2. Die Hypothese

Zunächst wird die Hypothese aufgestellt, dass die Effizienz des Informationssystems von der Menge der angebotenen Informationen abhängt.

Es wird angenommen, dass die Effizienz des Informationssystems von der Menge der angebotenen Informationen abhängt.

Es wird angenommen, dass die Effizienz des Informationssystems von der Menge der angebotenen Informationen abhängt.

2.2. INFORMATION-SYSTEM UND EFFIZIENZ

2.2.1. Überblick

In diesem Kapitel werden die Ergebnisse empirischer Studien zum Einfluß des Informationssystems auf die Effizienz komplexer betrieblicher Entscheidungen gewürdigt. Dabei werden zunächst die klassischen Aktionsparameter Menge und Aggregationsgrad der angebotenen Informationen dargestellt. Anschließend werden die Wirkungen der vornehmlich in jüngerer Zeit geprüften Variablen Medium und Präsentationsform untersucht. Das Kapitel wird abgeschlossen mit Gestaltungsgrößen eines aktiven Informationssystems: Feedback und formale Entscheidungshilfen.

Im Mittelpunkt unserer Bestandsaufnahme stehen Laborexperimente, in denen ökonomische Effizienzgrößen als abhängige Variable betrachtet werden, und zwar bei individuellen Entscheidungen. Diese Untersuchungen werden in der Regel als computergestützte Unternehmensspiele durchgeführt. Ob die dabei ermittelten Ergebnisse methodenspezifischer Natur sind, kann durch den Vergleich der gleichfalls ausgewerteten Feldstudien, die sich auf kollektive Entscheidungen beziehen, geprüft werden. Beim Parameter "Angebotsmenge" können zudem eine Reihe von Studien referiert werden, die in der Marketingforschung durchgeführt wurden.

2.2.2. Angebotsmenge und Effizienz

2.2.2.1. Die Hypothesen

Zum Einfluß der Angebotsmenge auf die Effizienz von Entscheidungen wurden zwei Hypothesen aufgestellt, die in partiellem Konflikt zueinander stehen:

Je größer die Menge an angebotenen Informationen, desto höher die Effizienz der Entscheidung.

Mit zunehmender Menge an angebotenen Informationen nimmt die Effizienz der Entscheidung zunächst zu und dann ab.

Die erste Hypothese verkörpert ein ungebrochenes Effizienzpostulat betrieblicher Informationssysteme. Sie geht von der recht mechanistischen Vorstellung aus, daß man nur die "richtigen" Bausteine anbieten müsse, um effiziente Entschlüsse zu bewirken. Diese Vorstellung wird heute nur noch selten vertreten. Sie hat in den sechziger Jahre eine beachtliche Rolle gespielt. Wie die Auswertung von Gemünden belegt, war das Streben nach "besserer", "schnellerer", "genauerer" Informationsversorgung der Geschäftsleitung ein zentrales Motiv für die Beschaffung von EDV-Anlagen und die Implementierung von Informationssystemen.¹⁾ In der politischen Diskussion, insbesondere über die Information von Verbrauchern, besitzt diese These auch heute noch einen erheblichen Stellenwert. Deshalb soll sie in unserer Bestandsaufnahme erneut zur Prüfung gestellt werden.

Die Alternativhypothese geht im aufsteigenden Ast davon aus, daß man ohne geeignete Informationen keine effiziente Entscheidung treffen könne. Sie unterscheidet sich jedoch vom ungebrochenen Effizienzpostulat dadurch, daß mit zunehmender Angebotsmenge die positiven Wirkungen ab- und die negativen Wirkungen zunehmen. Für eine nur degressive Zunahme des Informationsnutzens zusätzlich angebotener Informationen spricht vor allem die Überlegung, daß die zusätzlich angebotenen Informationen mit den bereits vorhandenen korreliert sind, so daß es zu einer Wissensduplizierung kommt. Zusätzliche Informationen würden dennoch benutzt, weil sie dem Streben der Entscheidungsträger nach höherer Urteilssicherheit und besserer Entschlußrechtfertigung entgegenkämen. Die Qualität ihrer Entscheidungen würde jedoch nur noch marginal verbessert.²⁾

Die Zunahme negativer Wirkungen wird damit begründet, daß die Bereitstellung zusätzlicher Informationsangebotsmengen mit mindestens linear steigenden Gesamtkosten verbunden ist. Zusätzlich rechnet man mit dysfunktionalen Verhaltenswirkungen: Die begrenzte Kapazität der Entscheidungsträger würde dazu verwendet, Informationen aufzunehmen und zu verwalten, statt sie problembezogen zu verarbeiten. Außerdem lenke die "Zusatz"-Information von zentralen Schlüsselinformationen ab. Lucas und Nielsen bringen dies auf die griffige Formel:

"Too much information forces the decision maker to spend time organizing the information instead of solving the problem. Irrelevant information distracts the decision maker and diverts attention from key variables."³⁾

Gegen die im fallenden Ast behauptete Ineffizienzvermutung wurde eingewendet, daß die betrieblichen Entscheidungsträger die auseinanderlaufenden positiven und negativen Effekte zusätzlicher Informationsmengen antizipieren, so daß sie nur den für sie relevanten Informationsbedarf decken.⁴⁾ Wie stichhaltig diese Anti-These ist, soll nun unsere Bestandsaufnahme der empirischen Befunde zeigen.

2.2.2.2. Effizienz betrieblicher Entscheidungen – jenseits einer kritischen Mindestmenge unabhängig von der Angebotsmenge

Eine Bestandsaufnahme von zehn empirischen Studien zeigt ein **uneinheitliches Ergebnis** (vgl. Abbildung 2.3.)⁵⁾:

- * bei fünf Studien ergeben sich keine signifikanten Unterschiede;
- * zwei Studien belegen einen signifikant positiven Einfluß der Angebotsmenge;
- * bei einer Studie zeigt sich ein signifikant ausgeprägter, umgekehrt u-förmiger Verlauf,
- * bei zwei Studien zeigt sich ein signifikanter negativer Einfluß der Angebotsmenge, in einer der beiden ist dieser Einfluß nur kurzfristig;
- * eine weitere Studie weist einen negativen Einfluß aus, macht aber keinerlei Aussagen über die Signifikanz des Effektes, nicht einmal über die Streuungen innerhalb der verglichenen Gruppen.

Läßt sich aus diesen heterogenen Befunden eine Entscheidung zwischen den (partiell) konkurrierenden Hypothesen ableiten? Wie stark sind die Einflüs-

Abbildung 2.3.: Einflüsse der Angebotsmenge auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Oskamp (Overconfidence 1965)*	Beurteilung eines psychisch Kranken	8 Ärzte 24 Studenten	Zusätzliche Informationsversorgung über den Patienten, 4 Lieferrunden	Anteil korrekt beantworteter Fragen je Runde	n.s. (trotz zunehmender Informationsversorgung bleibt Fehlerrate hoch)
Ebert (Decision Effectiveness 1972)	Unternehmensspiel, Abstimmung von Produktions- und Absatzmengen	60 Stud.	nur relevante vs. relevante und irrelevante (Feedback-) Informationen (zweite exp. Bed.: ohne/mit Planung)	Kosten (Rohwerte am Spielende: "terminal costs" laut "Table 3")	geringere Kosten, falls nur relevante Informationen, kein Signifikanztest, keine Effektschätzung wg. mangelnder Informationen
Witte (Informations-Verhalten 1972)	erstmalige Beschaffung von EDV-Anlagen, Feldstudie	227 Organisationen	Anzahl Informations-Versorgungs-Aktivitäten (relativiert an der Anzahl aller Aktivitäten eines Entscheidungsprozesses)	Innovationsgrad, Expertenurteil, berücksichtigt verschiedene Gütekriterien	n.s. (tendenziell umgekehrt u-förmige Verlaufsform), keine Effektschätzung wg. mangelnder Informationen
Bollom (Interim Reporting 1973)	Börsenspiel, stark vereinfacht	108 Stud.	ohne vs. mit zusätzliche Quartalsberichterstattung (fünf verschiedene Methoden)	Rendite	sign. höhere Rendite ohne Quartalsberichte (Re-Analyse von "Exhibit B", t-Test "U" vs. andere Gruppen, punktbiser. Korr.: -0,58)
McIntyre (Financial Statements 1973)	Aktienkauf	70 Stud. 5 Man.	Jahresabschlüsse nur mit historischen Kosten, nur mit Wiederbesch.-Kosten vs. mit beiden Kosteninformationen	Ob von 2 Aktien diejenige mit besserer Rendite gewählt wird	n.s. (Re-Analyse von "Table 6", "both" vs. andere Gruppen, tetrachor. Korrelation 0,09)
Benbasat (Effects 1974)	Unternehmensspiel, Beschaffung und Produktion	32 Stud.	Menge an abrufbaren Berichten (8 vs. 19); andere exp. Bed.: Tabelle vs. Graphik, ohne/mit Entscheidungshilfen, ohne vs. mit "exception reporting"	Kosten	n.s., auch keine sign. Interaktionseffekte, erkl. Varianz 1,7 v. H. (Re-Analyse von "Table 1", S. 105)
Grinyer/Norburn (Planning 1975)	Entscheidungen der Geschäftsführung bei strategischer Planung, Feldstudie	21 Unternehmen	Anzahl empfangener Informationen ("received informations") für Produkt- und Sortimentsentscheidungen	Rendite vor Zins und Steuern, im Vergleich zum jeweiligen Markt	n.s., Spearman'sche Rangkorrelation: 0,23
Lucas/Nielsen (Presentation 1980)	Unternehmensspiel zur Logistik eines Handelsunternehmens	36 Stud. 78 Manager	mit vs. ohne Zusatzinformation beim zweiten Spiel	erzielte Gewinne ("mean profit")	positiv, ohne Zusatzinformation keine sign. Verbesserung, mit Zusatzinformation sign. Verbesserung bei 5 von 7 Gruppen, keine Effektschätzung wg. mangelnder Informationen
Chorba/New (Information Support 1980)	Unternehmensspiel, Absatz, je vier Wettbewerber pro Spielgruppe	13 St. 12 Man. 2 Prof. 4 And.	teilweise vs. vollständige Informationslieferung (dritte Bed.: gebührenpflichtige Belieferung laut Nachfrage)	Rangplatz pro Gruppe bezgl. kumul. Einzahlungsübersch.	bei vollst. Information sign. besserer Rangplatz (Re-Analyse: U-Test auf Basis von "Table 4")
Ghani/Lusk (Information Representation 1982)	einfaches Unternehmensspiel, Absatzmengenentscheidung	49 Stud.	Zusätzliches Informationsangebot nach der vierten Spielperiode (zweite exp. Bed.: Tabelle vs. Graphik)	Gewinnveränderung	Kurzfristiger sign. Gewinnrückgang, gefolgt von sign. Anstieg auf altes Niveau
Hauschildt (Effizienz 1983)	Entscheidungen der Geschäftsführung, Feldstudie	73 Entscheidungen	Anzahl Informationen, die zum Nachfragezeitpunkt verfügbar sind	Entscheidungsqualität im ex-post-Urteil der Top-Manager	n.s., keine Effektschätzung wg. fehlender Informationen
Shields (Supply and Demand 1983)	Analyse von Kosten- und Leistungsberichten betr. Teileinheiten ("performance reports")	12 Manager je 4 Analysen	a) 3 vs. 9 zu beurteilende Teileinheiten b) Berichte mit 6 vs. 13 Beurteilungskriterien	Ob Analytiker die nach Expertenmeinung "kritische Ursache" herausfinden	a) n.s. b) n.s. sign. Interaktion: für gesamte Angebotsmenge umgekehrt u-förmiger Verlauf, erklärte Varianz 21,4 v. H. (Re-Analyse von "Table 9")

*: Die Studie von Oskamp wird nur nachrichtlich vermerkt, da sie sich nicht auf betriebliche Entscheidungen bezieht

se? Ist eine Variation der Angebotsmenge überhaupt beachtlich für die Effizienz der Entscheidung?

Betrachten wir zur Beantwortung dieser Fragen zunächst die fünf Studien **ohne** signifikante Unterschiede. Bei dreien kann man die Effektstärken entweder direkt ablesen oder mittels einer Re-analyse abschätzen. Es handelt sich um schwache Effekte: die mittlere erklärte Varianz liegt bei 2 v. H.⁶⁾ In den beiden übrigen Studien liegen hinreichend große Stichproben vor, um bei einem "wahren" Effekt mittlerer Stärke mit hoher Wahrscheinlichkeit eine signifikante Verwerfung der statistischen Nullhypothese zu ermöglichen.⁷⁾ Wir können demnach als erstes Zwischenergebnis festhalten: In der **Mehrheit der Untersuchungen** findet sich **kein deutlich ausgeprägter Zusammenhang** von Angebotsmenge und Effizienz der Entscheidung.

Doch: Eine "Menge" an empirischen Prüfungen ist nicht gleichzusetzen mit einer hohen "Strenge" der Prüfung im einzelnen Fall. Wir nehmen diesen berechtigten Einwand, der häufiger gegen Meta-Analysen vorgetragen wird,⁸⁾ ernst, und gehen etwas ausführlicher auf die besonders strenge Prüfung bei Benbasat ein.⁹⁾

Benbasat führt ein computergestütztes Unternehmensspiel zur Beschaffungs- und Produktionsplanung durch. An zehn verschiedenen Entscheidungszeitpunkten müssen die Bestellmengen und -zeitpunkte und die Produktionsmengen für jeweils 20 Spielperioden festgelegt werden. In dem Unternehmensspiel werden vier unabhängige Variablen experimentell manipuliert: Form der Datenpräsentation (Tabelle vs. Graphik), ohne vs. mit Entscheidungshilfen, ohne vs. mit "exception reporting", wenige vs. viele angebotene Berichte. Zwei weitere unabhängige Variablen sind die Persönlichkeitsmerkmale: Entscheidungsstil (Feld(un)abhängigkeit) und problembezogenes Vorwissen. Als abhängige Variablen werden die erzielten Kosten, der Zeitverbrauch für die Entscheidungen und die Anzahl nachgefragter Berichte verwendet.

Von besonderem Interesse ist an dieser Stelle die Operationalisierung der "Angebotsmenge". (Die Befunde und Operationalisierungen der anderen unabhängigen Variablen werden in späteren Kapiteln unserer Bestandsaufnahme referiert). Benbasat bildet zwei Faktorstufen: In der Vergleichsgruppe¹⁰⁾

("necessary group") werden acht Berichte angeboten, die Benbasat als "notwendige" Informationen ansieht, um eine optimale Entscheidung zu treffen. Die experimentelle Gruppe ("overload group") erhält elf weitere Berichte, die zwar problembezogene Informationen enthalten, zur Bestimmung einer optimalen Entscheidung aber nach Meinung von Benbasat nicht benötigt werden.

Seine Studie zeigt, daß die Gruppe mit der größeren Angebotsmenge signifikant mehr Berichte beschafft: In der Vergleichsgruppe werden pro Entscheidung durchschnittlich 6,43 Berichte abgerufen, in der experimentellen Gruppe sind es 9,11. Die erklärte Varianz ist mit 36,4 v. H. recht beachtlich.¹¹⁾ Hervorzuheben ist die Verteilung der Informationsabrufe auf "notwendige" und "zusätzlich" angebotene Berichte: Die experimentelle Gruppe ruft durchschnittlich 3,62 Zusatzberichte und 5,49 "notwendige" Berichte ab. Damit ergibt sich bei den als "notwendig" angesehenen Berichten eine **schlechtere** Informationsversorgung als bei der Vergleichsgruppe. Bezüglich des Zeitverbrauches ergeben sich keinerlei signifikante Haupt- oder Interaktionseffekte. Somit sind alle Voraussetzungen der Informationsüberlastungshypothese erfüllt:

- * Das zusätzliche Informationsangebot hat einen geringeren Informationsnutzen als das Basisangebot an notwendigen Informationen.
- * Die Entscheidungsträger nutzen die zusätzlich angebotenen Informationen.
- * Durch das zusätzliche Angebot schränken die Entscheidungsträger die Beschaffung notwendiger Kerninformationen ein.
- * Ein größerer Teil der insgesamt konstanten Bearbeitungszeit wird für Informationsbeschaffung und -verwaltung verwendet und damit der Informationsverarbeitung entzogen.
- * Weitere dysfunktionale Ablenkungseffekte können sich daraus ergeben, daß man Zusatzinformationen bei der Verarbeitung stärker gewichtet als Kerninformationen.

Trotzdem zeigt sich nicht der erwartete Informations-"Über"-Lastungs-Effekt. Betrachtet man die bei Benbasat ausgewiesenen Reports, dann ist die Falsifikation nicht auf eine grundlegend falsche Einschätzung der Qualität der angebotenen Daten zurückzuführen. Man muß vielmehr davon ausgehen, daß die Ursache für die Falsifikation der Hypothese auf einer zu mechanistischen Einschätzung des menschlichen Informationsverhaltens beruht. Informationen sind nur die Bausteine für einen komplexen Produktionsprozeß. Wie sie **verarbeitet** werden, wird nicht untersucht. Gutes, mengenmäßig angemessenes dosiertes Informations-"Material" ist demnach offenkundig nur eine notwendige, aber noch keine hinreichende Bedingung für eine gute Entscheidung. Es stellt sich daher für die zukünftige Forschung das Problem, menschliches Informationsverhalten differenzierter zu erforschen.

Eine systematische Bestandsaufnahme fordert, daß auch die übrigen Studien (mit signifikanten Einflüssen) berücksichtigt werden. Hierzu ist folgendes zu berichten:¹²⁾

1. In dem **Marketingspiel von Chorba und New** erhält ein "Wettbewerber 1" eine sehr geringe Informationsmenge (5 Items). Der "Wettbewerber 4" erhält die maximale Informationsmenge des Spiels (92 Items). Bei derart krassen Unterschieden verwundert es nicht, daß "Wettbewerber 4" signifikant besser abschneidet.¹³⁾ Man muß wohl davon ausgehen, daß der "Wettbewerber 1" nur einen Bruchteil der problemrelevanten Information erhielt und unter "information underload" litt.

Interessanter ist das Abschneiden der Wettbewerber 2 und 3. Sie konnten gegen Gebühr so viele Informationen kaufen, wie sie wünschten (maximal alle 92 Items). Tatsächlich beschafften sie im Durchschnitt aber nur einen kleinen Teil der angebotenen Informationen, nämlich 21,7 Items (Standardabweichung 18,0 Items). Mit dieser Informationsmenge erreichten sie im Durchschnitt einen Effizienzwert, der signifikant über dem von "Wettbewerber 1" und knapp unter dem von "Wettbewerber 4" liegt.¹⁴⁾ Unterstellt man, daß "Wettbewerber 4" mit dem kompletten, kostenfreien Informationsangebot mehr Informationen nutzte als Wettbewerber 2 und 3, dann ist jenseits der beschafften 21,7 Items kein Effekt der Informationsmenge mehr festzustellen.

Insgesamt gesehen bestätigt die Studie von Chorba und New die Hypothese, daß die Effizienz der Entscheidung mit der Menge an problemrelevanten Informationen **zunächst** ansteigt. Jenseits einer recht niedrig anzusetzenden Mindestmenge besteht kein enger Zusammenhang mehr.

2. In dem **logistischen Unternehmensspiel von Lucas und Nielsen** wird geprüft, wie sich eine **Informationsbeschränkung** auswirkt. In einem ersten Lauf des Unternehmensspiels erhalten alle sieben Versuchsgruppen und die Kontrollgruppe nur eine Teilmenge der üblichen Informationen ("four basic reports"). Im zweiten Lauf erhalten alle sieben Versuchsgruppen die volle Informationsmenge ("all reports"), während die Kontrollgruppe weiterhin nur die genannte Teilmenge des Angebotes erhält. Die Autoren unterstellen, daß das volle, übliche Informationsangebot zwar reichhaltig sei, aber nicht zu Überlastungseffekten führe. Im Ergebnis zeigt sich, daß fünf der sieben Versuchsgruppen im zweiten Spiel signifikant höhere Gewinne erzielen als im ersten. Bei der Kontrollgruppe ergeben sich zwar auch höhere Gewinne, aber der Zuwachs ist nicht signifikant. Doch die Kontrollgruppe besteht wie die beiden Versuchsgruppen **ohne** signifikante Lernfortschritte aus studentischen Versuchspersonen, so daß man den Befund auch mit **populationsspezifischen Differenzen** erklären kann. Alle drei studentischen Versuchsgruppen schneiden nämlich beim ersten Versuch besser ab als alle fünf Praktikergruppen, im zweiten Versuch liegen sie eher unter den Praktikergruppen. Außerdem ist die konkurrierende Hypothese zu beachten, daß nicht die Vergrößerung des Informationsangebotes, sondern **Lernfortschritte** bei dessen Nutzung die Effizienzsteigerung bewirkt haben. Hierfür spricht die Beobachtung, daß die studentische Kontrollgruppe praktisch die gleiche Ergebnisverbesserung aufweist wie die beiden studentischen Versuchsgruppen.

Das Experiment von Lucas und Nielsen kann somit nicht als Bestätigung der Hypothese gewertet werden, daß eine Beschränkung auf eine Teilmenge der problemrelevanten Informationen zu einer Effizienzminderung führt.

3. In dem stark vereinfachten **Börsenspiel von Bollom** stellt sich heraus, daß die in verschiedenen Varianten angebotenen Quartalsberichte von den studentischen Versuchspersonen nicht sachgerecht genutzt werden. Sie hätten insbesondere in der Versuchsbedingung "früher steigender/heute fallender

Einkommenstrend" früher verkaufen müssen. Gerade diese, von den Zwischenberichten besser angezeigte Information, wurde jedoch nicht genutzt. Es handelte sich also um eine subjektiv gesehen irrelevante Information, deren Beschaffung und Verarbeitung zwar Kapazität verschlang, aber keinen Nutzen für den ungeübten Entscheidungsträger brachte. Der recht starke negative Effekt¹⁵⁾ beruht demnach nicht auf einer zu großen Menge an angebotenen Informationen, sondern auf **mangelndem Sachverstand** der verwendeten studentischen Analytiker. Das Experiment von Bollom liefert somit keine Bestätigung des "fallendes Astes" der umgekehrt u-förmigen Verlaufshypothese.

4. **Ghani und Lusk** berichten, daß die Teilnehmer die zusätzliche Information zunächst in einem spielerischen "trial-and-error-mode" genutzt hätten. Da in ihrem Spiel lediglich eine Differenzierung der Wetterprognose von zwei auf drei Zustände vorgenommen wird, kann der vorübergehende negative Effekt auch in dieser Untersuchung nicht auf die Menge, sondern allenfalls auf die Art der Information zurückgeführt werden.

5. **Shields** läßt zwölf Praktiker **Kosten- und Leistungsberichte** ("performance reports") auf mögliche Ursachen von Unwirtschaftlichkeiten hin analysieren. Jeder Praktiker untersucht die Berichte von vier Betrieben, die sich in der Zahl der zu beurteilenden betrieblichen Teileinheiten (3 vs. 9) und der zu berücksichtigenden Beurteilungskriterien (6 vs. 13) unterscheiden. Die Effizienz wird daran gemessen, ob die von Shields für wesentlich angesehene "Ursache" aus den möglichen Kategorien: Management, Arbeiter, Maschinen, Lagerhaltung, Controlling und "Andere" gefunden wird. Der Befund zeigt eine umgekehrt u-förmige Beziehung zwischen der Menge an angebotenen Informationen und dem Anteil angemessener Beurteilungen bei den einzelnen Versuchsbedingungen. Die erklärte Varianz der intra-individuellen Leistungsunterschiede ist mit 21,4 v. H.¹⁶⁾ recht beachtlich.

Die Untersuchung von Shields weist zwei Besonderheiten auf: Es wird ein Beurteilungsproblem mit einem punktuellen Effizienzmaß verwendet. Die Effizienzprüfung bezieht sich auf intra-individuelle Leistungsunterschiede bei verschiedenen Problemen, nicht auf inter-personelle Leistungsunterschiede bei derselben Entscheidung. Durch sein Meßwiederholungsdesign kann Shields die subjektabhängigen Einflußgrößen auf die Kapazität und die Informations-

verarbeitungsstrategie konstant halten und damit die Wirkungen der experimentell manipulierten Angebotsmenge schärfer testen.

Zur Studie von Shields ist kritisch anzumerken, daß sein punktuelles Effizienzmaß recht grob ist. Es berücksichtigt weder, wie differenziert die Ursachenforschung erfolgt, noch, welche Qualität die Initiativen besitzen, die ein Controller aufgrund seiner Analyse auslöst. Bezüglich des aus der Marketingforschung entlehnten Designs ist kritisch anzumerken, daß es einen Unterschied macht, ob man aus mehreren Abteilungen die "beste" oder "schlechteste" auswählen will, oder ob man für jede einzelne eine differenzierte Lagebeurteilung und Maßnahmenplanung vornehmen möchte.

Ob dies der Grund für den festgestellten umgekehrt u-förmigen Verlauf ist, kann ein Vergleich mit Experimenten aus der Marketingforschung zeigen. Dort wird in ähnlicher Weise geprüft, ob Konsumenten eine optimale Marke auswählen, wenn man die Angebotsmenge variiert. Diesen Studien wollen wir uns nun zuwenden.

2.2.2.3. Effizienz von Kaufentscheidungen: divergierende Befunde zum Einfluß von Alternativen- und Attributmengen

Die empirischen Wirkungsanalysen in der Marketingforschung widmen sich der Frage, ob "mehr" Informationen zu "besseren" Entscheidungen führen. Bekannt geworden sind die von Jacoby und seinen Mitarbeitern durchgeführten Untersuchungen. In ihnen soll geprüft werden, wie sich eine Veränderung der Anzahl angebotener Alternativen und Attribute auf die subjektiv optimale Wahl eines Konsumenten, die bei den jeweils angebotenen Marken möglich ist, auswirkt. Die Untersuchungen sind Gegenstand heftiger Kontroversen. Gestritten wurde u. a. über die Angemessenheit der vorgelegten Befundinterpretationen, der verwendeten Effizienzmaße und der Gleichsetzung der unabhängigen Variablen Alternativen- und Attributmengen.¹⁷⁾

Hervorzuheben ist die Re-Analyse mehrerer Studien durch Malhotra, Jain und Lagakos.¹⁸⁾ Sie kommen aufgrund einer Logitanalyse zu anderen Schlußfolgerungen als die Autoren. Obwohl man die Logitanalyse als angemessenes Ver-

fahren für diese Fragestellung bezeichnen muß, ist ihre Re-Analyse praktisch wertlos. Sie leidet vor allem darunter, daß sie – entgegen ihrem eigenen Anspruch – keine theoretisch abgeleitete Funktion testen, sondern stattdessen mit sehr vielen Dummy-Variablen arbeiten und jede Studie für sich re-analysieren. Dieses führt zu einem ungünstigen Verhältnis von Parameterzahl und Beobachtungen. Das Ergebnis sind zahlreiche nicht-signifikante Parameter mit sehr großen Standardabweichungen.

Aus diesem Grunde haben wir eine eigene Re-Analyse durchgeführt. In unseren Logit-Regressionen werden die Anzahlen der angebotenen Marken und Alternativen als quantitative, metrische Variablen und nicht als "Dummies" behandelt. Geschätzt werden sowohl lineare als auch quadratische Regressionen. Beide Modelle werden mit und ohne Interaktionen geschätzt.¹⁹⁾ Außerdem haben wir getestet, ob es ein gemeinsames Modell mit einem **einheitlichen** Kurvenverlauf gibt, daß alle durchgeführten Studien von Jacoby et al. und Malhotra erklärt.

Auf die Ergebnisse dieses Versuches wird weiter unten eingegangen. Zunächst sind die Resultate der einzelnen Studien zu referieren. Wir gliedern die Zusammenfassung der Ergebnisse nach den unabhängigen Variablen: Anzahl angebotener Alternativen, Anzahl angebotener Marken und Anzahl insgesamt angebotener Informationseinheiten ("Gesamtverlauf"), da behauptet wurde, daß es mit zunehmender Alternativenzahl schwieriger werde, die subjektiv optimale Lösung zu finden, während eine zunehmende Attributzahl die Wahl eher erleichtere.²⁰⁾

Es ergibt sich folgendes Ergebnis (vgl. Abbildung 2.4):

a) Anzahl angebotener Marken ("Alternativenzahl"):

Der Einfluß ist **uneinheitlich**: Zwei Studien zeigen einen signifikant positiven Einfluß, eine Studie zeigt einen signifikant negativen Einfluß, zwei Studien zeigen keine signifikanten Unterschiede.²¹⁾

Abbildung 2.4.: Einflüsse der Angebotsmenge auf die Effizienz von Kaufentscheidungen von Konsumenten

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Jacoby/Speller/Kohn (Information Load 1974)	Beschaffung von Waschmitteln ("laundry detergent")	153 Stud.	a) 4 vs. 8 vs. 12 angebotene Marken b) 2 vs. 4 vs. 6 angebotene Attribute pro Marke	Ob subjektiv beste Marke gewählt wird oder nicht	a) positiv sign. b) positiv sign. sign. Interaktionseffekt, für gesamte Angebotsmenge umgekehrt u-förmiger Verlauf (Re-Analyse: Logitanalyse mit GLIM)
Jacoby/Speller/Kohn-Berning (Replication 1974), 1. Exp.	Beschaffung von Reisprodukten ("instant rice")	192 Hausfrauen	a) 4 vs. 8 vs. 12 vs. 16 angebotene Marken b) 4 vs. 8 vs. 12 vs. 16 angebotene Attribute pro Marke	1. Ob subjektiv beste Marke gewählt wird 2. Rangkorrelation zwischen Einstufung nach Exp. und Nutzenrang aufgrund vorheriger Frage	1. Gewählte Marke: a) n.s. (positiv) b) negativ tend. sign. Interaktion: n.s. (Re-Analyse: Logitanalyse mit GLIM) 2. Rangkorrelation: a) n.s. b) positiv sign. Interaktion: n.s.
Jacoby/Speller/Kohn-Berning (Replication 1974), 2. Exp.	Beschaffung von Fertiggerichten ("prepared dinners")	192 Hausfrauen	a) 4 vs. 8 vs. 12 vs. 16 angebotene Marken b) 4 vs. 8 vs. 12 vs. 16 angebotene Attribute pro Marke	1. Ob subjektiv beste Marke gewählt wird 2. Rangkorrelation zwischen Einstufung nach Exp. und Nutzenrang aufgrund vorheriger Frage	1. Gewählte Marke: a) positiv sign. b) n.s. (positiv) Interaktion: negativ sign. (Re-Analyse Logitanalyse mit GLIM) 2. Rangkorrelation: a) n.s. b) positiv sign. Interaktion: n.s.
Scammon ("Information" Load 1977)	Beschaffung von Erdnußbutter	297 "Kali-for-nier"	Kurzfilme (30 Sek.) zwei Produkte ohne Nährstoffangaben vs. 4 vs. 8 Angaben; (zweiter exp. Faktor: numerische vs. verbale Angaben)	Ob das "objektiv" bessere Produkt "Koogle" besser bewertet wird	n.s. keine sign. Unterschiede zwischen allen drei Gruppen, "ohne" jedoch sign. schlechter als "mit" Nährstoff- und Kalorienangaben; (Re-Analyse von "Table 3" tetrach. Korr.: 0,38; kein Interaktionstest)
Sproles/Geistfeld/Badenhop (Efficient Consumers 1980) 1. Exp.	Beschaffung von Decken ("blankets")	141 Stud.	4 Produkte ohne jegl. Zusatzinformation vs. zusätzlich handelsübl. Verpackungsinf. vs. zusätzlich Testinf. Markennamen bei keiner Bed. bekanntgegeben	Ob das laut Warentest beste Produkt gewählt wird	positiv sign., keine Effektschätzung wg. mangelnder Informationen

Abbildung 2.4.: Einflüsse der Angebotsmenge auf die Effizienz von Kaufentscheidungen von Konsumenten

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Sproles/Geistfeld/Badenhop (Efficient Consumers 1980) 2. Exp.	Beschaffung von Kochern ("slow cooker")	142 Stud.	4 Produkte ohne jegl. Zusatzinformation vs. zusätzlich handelsübl. Verpackungsinf. vs. zusätzlich Testinf. Markennamen bei keiner Bed. bekanntgegeben	Ob das laut Warentest beste Produkt gewählt wird	n. s., keine Effektschätzung wg. mangelnder Informationen
Henry (Processing Accuracy 1981)	Beschaffung von Waschmitteln ("laundry detergent")	28 Frauen je 4 Entsch.	Anzahl angebotener Attribute pro Marke: 2,3,4,5 sequentiell pro Vpn variiert, Anzahl Marken jeweils 11; außerdem wird die kognitive Komplexität (Paragraph Completion Test) regressionsanalytisch konstant gehalten	Diskriminierungsleistung: Übereinstimmung der wahrgenommenen und "objektiven" paarweisen Ähnlichkeiten der Marken	tend. sign. negativ, (bei zweiseitigem Test, t-Wert des Regressionskoef.: -1,74, entspräche bei univariatem Test punkt-biser. Korrr. von -0,33)
Malhotra (Information Load 1982)	Beschaffung von Häusern	300 Haushaltsvorstände	a) 5 vs. 10 vs. 15 vs. 20 vs. 25 angebotene Alternativen b) 5 vs. 10 vs. 15 vs. 20 vs. 25 angebotene Attribute pro Alternative	1. Ob subjektiv beste Alternative gewählt wird 2. Ob subjektiv wenigstens zweitbeste Alternative gewählt wird	1. Beste Alternative: a) negativ sign. b) negativ sign. Interaktion: n.s. (Re-Analyse: Logitanalyse mit GLIM) 2. Zweitbeste Altern.: a) negativ sign. b) negativ sign. Interaktion: n.s.
Berndt (Informationsmenge 1984)	Beschaffung von Marmelade	154 Stud.	a) 3 vs. 5 vs. 7 angebotene Marken b) 3 vs. 5 vs. 7 angebotene Attribute pro Marke	Übereinstimmung ex-ante und ex-post erfragter Relevanz der drei wichtigsten Attribute	a) n.s. b) n.s. (kein Test auf Interaktion)
Silberer/Förster/Schwetz (Wirkungen 1984), 1. Exp	Beschaffung von Mono-Radio-Recordern	100 Stud. 100 Schüler 80 Verbraucher	a) ausführliche vs. kurzgefaßte Testberichte b) ohne vs. mit Gesamtnoten	Rangkorrr. Alternativen "engere Wahl" ex-ante mit ex-post	a) n.s. b) n.s. (kein sign. Interaktionseffekt)
Silberer/Förster/Schwetz (Wirkungen 1984), 2. Exp.	Beschaffung von Mono-Radio-Recordern	98 Schüler	a) ausführliche vs. kurzgefaßte Testberichte b) eine Gesamtnote vs. mehrere alternative Gesamtnoten pro Produkt	Rangkorrr. Alternativen "engere Wahl" ex-ante mit ex-post Urteil	a) n.s. b) n.s. (Interaktion: n.s.) (ohne Telefunken-Gerät, gem. Autorenempfehl.).

b) Anzahl angebotener Attribute ("Kriterienzahl"):

Der Einfluß ist **uneinheitlich**: Zwei Studien zeigen einen signifikant positiven Einfluß, zwei Studien zeigen einen signifikant negativen Einfluß, eine Studie einen tendenziell signifikanten, negativen Einfluß, vier Studien zeigen keine signifikanten Unterschiede.²²⁾

c) Anzahl aller angebotenen Informationseinheiten ("Gesamtverlauf"):

Der Einfluß ist **uneinheitlich**: Zwei Studien zeigen einen signifikant ausgeprägten umgekehrt u-förmigen Verlauf, eine Studie einen signifikant negativen Einfluß, zwei Studien zeigen keine signifikanten Unterschiede.²³⁾

In zwei weiteren Experimenten von Silberer, Förster und Schwetz wird die Wirkung von "ausführlichen" vs. "kurzgefaßten" Testberichte untersucht. Als zweiter Faktor wurde die Aggregationsvariable²⁴⁾ "ohne" vs. "mit" sowie "eine" vs. "mehrere" Gesamtnoten variiert. Bei keiner der Variablen ergibt sich ein signifikanter Haupt- oder Interaktionseffekt.

Faßt man alle Ergebnisse zusammen, so zeigt sich bei 10 der 21 untersuchten Variablenbeziehungen kein signifikanter Zusammenhang. Da die Stichprobenumfänge meistens deutlich über 100 liegen, **kann man nicht davon ausgehen, daß zwischen der Menge an angebotenen Informationen und der Effizienz einer Kaufentscheidung ein beachtlicher, generell zu erwartender Zusammenhang besteht.** Dieses Ergebnis gilt unabhängig davon, ob man Studien betrachtet, die als Effizienzmaß eine nach Expertenmeinung beste Wahl oder eine laut individueller Gewichtung günstigste Entscheidungsmöglichkeit zugrundelegen.

Daß man das Ergebnis nicht auf mangelnde Vergleichbarkeit der Studien zurückführen kann, zeigt vor allem eine Re-Analyse der vom Design und vom Effizienzmaß her ähnlich angelegten Studien von Jacoby et al. und Malhotra. Unser Versuch, eine einheitliche Funktion zu schätzen, die alle vier Experimente der Autoren erklärt, muß als gescheitert angesehen werden. Unsere diesbezüglichen Tests zeigen einen signifikanten Zuwachs der bei GLIM ausgewiesenen "deviance", wenn man unterschiedliche Verlaufsformen bei den einzelnen Studien zuläßt. Dieses Ergebnis entspricht den Daten, die sehr

verschiedenartige Verläufe bei den einzelnen Experimenten zeigen.

Wir glauben nicht, daß man die konfliktären Ergebnisse durch eine angemessenere Datenanalyse auflösen kann. So ergibt unsere Re-Analyse der Daten der ersten Studie von Jacoby, Speller und Kohn ein lineares Modell mit Interaktionseffekt, das nur drei Parameter benötigt. Alle drei Parameter sind signifikant. Seine Anpassung an die Daten ist nicht signifikant schlechter als die des Dummy-Modells von Malhotra, Jain und Lagakos mit sehr viel mehr Parametern. Insofern ist die Datenanalyse angemessener als bei den bisherigen Primär- und Sekundäranalysen. Trotzdem handelt es sich um ein singuläres, stichprobenabhängiges Ergebnis, auch wenn es im Konflikt zu den bisherigen Interpretationen steht.²⁵⁾ Was fehlt, ist eine **Theorie**, die den Test leitet und die Ergebnisse **aller** Studien erklärt. Die statistische Datenanalyse kann immer nur die Implikationen eines gewählten Modelles aufzeigen, das aufgrund einer theoretischen Vermutung aufgestellt wurde. Ersetzen kann die Datenanalyse die Theorie nicht – auch dann nicht, wenn man mit der Logitanalyse ein "neuartiges" Verfahren anwendet, das der Datenstruktur besser angemessen ist.

Da die Studien aus dem Marketingbereich für unser Thema nur eine mittelbare Relevanz besitzen und sich auch bei gut vergleichbaren Experimenten heterogene Befunde ergeben, wollen wir darauf verzichten, eine Ursachenanalyse der Divergenzen mittels einer detaillierten Würdigung der Einzelstudien vorzunehmen. Wir beschränken uns vielmehr auf einige generelle Anmerkungen:

1. Die Arbeiten von Henry, Scammon und Sproles et al. besitzen unseres Erachtens eine besonders geringe Validität, weil sie von einem "objektiven" Urteil ausgehen, das sich mit der Gewichtung der Entscheidungskriterien, die der einzelne Konsument zugrundelegt, keineswegs decken muß.²⁶⁾

2. In den übrigen Studien wird gemessen, ob der Entscheidungsträger eine Wahl trifft, die nach seiner Gewichtung als optimal anzusehen ist. So ermitteln beispielsweise Jacoby et al., ob der Konsument die Marke auswählt, die in der jeweils angebotenen Alternativenmenge am nächsten bei seinem Idealprodukt liegt.²⁷⁾ Auch diese Konzeption geht von vereinfachenden Prämissen aus:

- * Sie unterstellt ein **unbeschränktes "Optimierungs"-Streben** der Entscheidungsträger: Wenn es ein Produkt gibt, das dem Konsumenten einen höheren Nutzen stiftet, dann soll er es auch wählen, und zwar unabhängig davon, wie hoch der (erwartete) zusätzliche Informationsaufwand im Verhältnis zum (erwarteten) Zusatznutzen ist. Dies ist nicht informations-ökonomisch.
- * Das Effizienzmaß geht von einer **falsch spezifizierten "kognitiven Algebra"** aus. Es unterstellt eine rein kompensatorische Bewertung. Es gibt jedoch eine ganze Reihe von Untersuchungen, die belegen, daß sich die Informationssuche und -verarbeitung mit der Komplexität des Problems verändert: Je komplexer die Entscheidung ist, desto eher werden Eliminationsregeln angewandt, die sich auf eine Teilmenge der Informationen stützen. Die unterstellte kompensatorische Bewertung findet häufig erst in einer "zweiten Phase" des Entscheidungsprozesses statt.²⁸⁾
- * Die **Relativierung** der getroffenen Wahl an den Nutzenwerten der jeweils **angebotenen Alternativen** ist **problematisch**. Hierzu ein fiktives Beispiel: Bei einer 3-Alternativen-Situation könne der Konsument die subjektiven Nutzenwerte 5, 8 und 10 realisieren, bei einer 5-Alternativen-Situation seien die Werte 3, 6, 12, 24, 30 möglich. Wenn sich der Konsument in der zweiten Situation für die Alternative mit dem Nutzenwert 24 entscheidet, dann hat er nach Jacoby et al. "relativ" gesehen, schlechter gewählt, als wenn er in der ersten Situation die Alternative mit dem Nutzenwert 10 ergreift. Absolut betrachtet hat er sich jedoch erheblich verbessert. Warum soll man diese Verbesserung gerade an dem erreichbaren Maximum "gegebener" Alternativen relativieren und nicht an dem tatsächlichen individuellen Entscheidungsaufwand des gerade betrachteten Entscheidungsträgers? Wie ist zu verfahren, wenn man die unrealistische Prämisse des Experimentes aufhebt, daß die Alternativen "gegeben" seien und stattdessen zuläßt, daß man sie aktiv suchen, verhandeln oder selbst entwickeln kann?

2. Schröder, Driver und Straußfeld vergleichen in ihren Experimenten Versuchssubjekten, denen sequentiell und gleichzeitig unterschiedlich viele Informationsmöglichkeiten angeboten werden.²⁹⁾ In den Experimenten aus der Markt-

* Die Effizienzkonzeption geht von der problematischen Prämisse aus, daß das Wissen über die **Zieleigenschaften und ihre Gewichte durch den Informationsprozeß nicht verändert** würden. Sie unterstellt, daß die Entscheidungsträger bereits vor dem Informationsprozeß wissen, was das für sie ideale Produkt sei. Die Ergebnisse der empirischen Zielforschung bei innovativen Entscheidungen belegen, daß man erst im Verlauf von Entscheidungs- und Informationsprozessen lernt, was man alles wollen kann, was die Entscheidungskriterien bedeuten und in welchen Ausprägungen sie empirisch auftreten.²⁹⁾ **Informationsprozesse sind Lernprozesse.** Wer die Effizienz an Kriterien und Gewichten mißt, die vor dem Prozeß geäußert wurden, der vernachlässigt diese Lernprozesse.

* Wer eine simple Präferenzstruktur hat, die sich auf sehr wenige Attribute stützt und diese in einer sehr einfachen Weise verknüpft, hat es sehr viel leichter, eine für ihn "optimale" Alternative zu finden. Eine höhere kognitive Leistung des stärker Differenzierenden wird durch die Effizienzkonzeption im Zweifel bestraft.

Die Ableitungen zur Informations-"Überlastungs"-Hypothese stützen sich auf zwei Grundpfeiler: die These von der begrenzten menschlichen Informationsverarbeitungskapazität und die Untersuchungen von Schroder, Driver und Streufert. Beide Grundpfeiler tragen nicht:

1. Millers "magical number seven - plus or minus two" bezieht sich auf die begrenzte Kapazität des **Kurzzeit-Gedächtnisses**.³⁰⁾ Es ist jedoch unzulässig, aus dieser begrenzten Gedächtnisspanne - selbst wenn sie für alle Menschen tatsächlich in gleicher Weise gilt - auf eine bei allen Menschen gleich begrenzte Problemlösefähigkeit zu schließen: Was gespeichert wird, wie sich die Integration" zu den sog. "chunks" vollzieht, hängt von den bei Anfängern und Experten außerordentlich unterschiedlich ausgeprägten Langzeitgedächtnissen ab, um nur eine inter-individuell oder inter-temporal unterschiedlich ausgeprägte Determinante ins Spiel zu bringen.³¹⁾

2. Schroder, Driver und Streufert vergleichen in ihren Experimenten Versuchsgruppen, denen **sequentiell** pro Zeiteinheit unterschiedlich viele Informationseinheiten angeboten werden.³²⁾ In den Experimenten aus der Marke-

tingforschung wird **simultan** eine bestimmte Angebotsmenge offeriert. Der Entscheidungsträger hat also vielmehr Freiheitsgrade, ob, wann und welche Information er aufnehmen und verarbeiten möchte. Es trifft zwar zu, daß die Beschaffungsmenge in Experimenten mit Information-Display-Boards stets (signifikant) zunimmt, aber dieser Zuwachs verläuft unter-proportional zur Angebotsmenge. Die Ursache kann darin gesehen werden, daß die Entscheidungsträger **Vereinfachungsstrategien** anwenden, die es ihnen ermöglichen, mehr oder weniger systematisch auf die Aufnahme und Verarbeitung bestimmter Informationen zu verzichten. Solche Heuristiken hängen von der Art der Informationspräsentation (sequentiell vs. simultan) und von der Art des Problems ab. Bei Schroder, Driver und Streufert geht es nicht um ein **Auswahl**-Problem, eine gegebene, optimale Alternative zu suchen. In ihrem militärischen Strategiespiel sind vielmehr unterschiedliche Informationen zu einer Lagebeurteilung zu integrieren und selbst Maßnahmen zu entwickeln. Deshalb kann man sich bei der Hypothesenableitung auch nicht auf die Befunde (!) von Schroder et al. beziehen. Darüber hinaus ist zu beachten, daß Schroder, Driver und Streufert eine Maßgröße des Informationsprozesses, nämlich Zahl und Komplexität der Integrationen und nicht etwa eine punktuelle Ergebnisgröße als abhängige Variable verwenden.³³⁾

Insgesamt gesehen sind die Untersuchungen zur sog. Informations-"Über"-Lastungs-Hypothese zu Recht kontrovers diskutiert worden. Unseres Erachtens sind zwei (miteinander verknüpfte) Defizite besonders gravierend:

1. **Theoriedefizit:** Man schließt aus dem Verlauf eines problematischen Effizienzmaßes – gestützt auf einige vage Analogien zu Miller und Schroder, Driver und Streufert – daß Informations-"Über"-Lastung vorgelegen haben müsse.

2. **Methodendefizit:** Es wird nicht empirisch kontrolliert, welche Informationen tatsächlich aufgenommen und verarbeitet werden.

Beides zusammen ergibt eine Forschung, die unseres Erachtens das Prädikat "verhaltens"-orientiert noch nicht verdient. Wenn es eines Beleges bedurfte, daß man menschliches Informations-Verhalten differenzierter erforschen sollte, so hat ihn diese Forschung erbracht. Zu befürchten ist allerdings,

daß die kontroverse Diskussion der Sache gelegentlich mehr geschadet als genutzt hat und die Kräfte auf das falsche Problem lenkte.

2.2.2.4. Wirkungen der Angebotsmenge – ein Zwischenergebnis

Faßt man die bisher referierten Studien zusammen, so lassen sich nur zwei ziemlich triviale Ergebnisse absichern:

- * Unterhalb einer vermutlich recht niedrig anzusetzenden Angebotsmenge an problemrelevanten Informationen ist keine gute Entscheidung möglich. Diese These wird vor allem durch die Studie von Chorba und New gestützt.
- * Bei einer sehr großen Angebotsmenge, d. h. jenseits von 10 Attributen und 5 Alternativen ist nicht damit zu rechnen, daß ein Entscheidungsträger ohne formale Entscheidungshilfen eine Wahl trifft, die nach seinen vor dem Informationsprozeß festgelegten Kriteriengewichten als produkt-nutzenmaximierend anzusehen ist. Diese These wird durch die Studie von Malhotra erhärtet.

Zwischen diesen beiden Extremen liegt eine sehr breite Grauzone, in der in den meisten Fällen kein signifikanter Zusammenhang zwischen der Angebotsmenge und der durchschnittlich erreichten Effizienz der Entscheidung nachgewiesen wird.

Wir führen dies darauf zurück, daß die einzelnen Entscheidungsträger bei unterschiedlichen Entscheidungsproblemen die Informationen in sehr unterschiedlicher Weise aufnehmen und verarbeiten, so daß das Verhältnis von "erklärter" zu "unerklärter" Varianz bedeutungslos niedrig ausfällt. Aufgrund einer reinen Variation der Angebotsmenge sind keine brauchbaren Prognosen über die mittlere Effizienz möglich. Man muß deshalb nach anderen System- oder Persönlichkeitsmerkmalen suchen, die eine bessere Effizienzprognose ermöglichen oder durch eine differenzierte Verhaltensanalyse klären, warum eine Variation der Angebotsmenge im Durchschnitt gesehen keinen Einfluß zeigt.

Gestaltungsempfehlungen sollte man aufgrund der bisherigen Forschung nicht ableiten. Die Konstrukteure und Nutzer von Informationssystemen könnten jedoch prüfen, ob eine Verringerung der Angebotsmenge Kostenvorteile bietet, ohne daß sich die Entscheidungsqualität merklich verschlechtert. Dies könnte vor allem dann der Fall sein, wenn:

- * der Wert der gespeicherten Informationen überschätzt wird,
- * die Nutzer vornehmlich nach Konfidenz streben und die Informationen in erster Linie entschluß-rechtfertigend einsetzen,
- * die Nutzer den Zeitaufwand für die Verarbeitung der angebotenen Daten unter- bzw. ihre eigene Kapazität überschätzen.

In der Untersuchung von Chorba und New zeigte sich, daß die Nutzer ihre Informationsbedürfnisse problemgerechter definierten, wenn die Informationsbeschaffung gebührenpflichtig war und der Nutzen der Informationsverwertung ihnen zugerechnet wurde.³⁴⁾ Dies könnte eine Option für die dynamische Anpassung von Informationsangebot, Informationsbedarf und Informationsnachfrage sein.

2.2.3. Aggregationsgrad und Effizienz

2.2.3.1. Die Hypothese

Das Augenmerk der Forschung zur Aggregation richtet sich auf die **Qualität**, nicht auf die Menge der angebotenen Daten. Man geht davon aus, daß man Entscheidungen verbessern kann, wenn man mit weniger Daten die wesentliche **Information** vermittelt. Die Effizienzbehauptung stützt sich auf eine Verbesserung des Nutzen-Kosten-Verhältnisses: Mit einem geringeren Aufwand für den Entscheidungsprozeß kann ein wenigstens gleich guter Entschluß getroffen werden. Wir konzentrieren uns auf die Nutzendimension, d. h. auf die Wirkungen des Aggregationsgrades auf die **Qualität** der Entscheidung. Auf die behauptete Verringerung des Entscheidungsaufwandes werden wir bei der Diskussion der Forschungsergebnisse zurückkommen, und zwar nach der Bestands-

aufnahme der Ergebnisse zur Qualitäts-Hypothese.

Die Hypothesen zum Zusammenhang von "Aggregationsgrad" und Entscheidungsqualität laufen darauf hinaus, einen **umgekehrt u-förmigen Verlauf** zu postulieren.³⁵⁾ Dabei können drei Varianten der Hypothesenableitung unterschieden werden:

a) Aggregierte statt disaggregierter Daten:

Bei dieser Hypothesenvariante geht man davon aus, daß mit zunehmender Verdichtung der Daten die entscheidungsrelevante Information zunächst erhalten bleibt bzw. erhalten werden kann. Es soll sich jedoch eine Verbesserung der Entscheidungsqualität einstellen, weil der Entscheidungsträger auf die wesentliche Information gelenkt wird und diese auch vollständig verarbeiten kann. Mit weiter zunehmender Informationsverdichtung kann die entscheidungsrelevante Information jedoch nicht mehr erhalten werden. Wegen dieses Informationsverlustes rechnet man mit Effizienzeinbußen.

b) Disaggregierte Daten zusätzlich zu aggregierten Daten:

Bei dieser Hypothesenvariante geht man vom umgekehrten Pol der Aggregationsskala aus. Man fragt: Was passiert, wenn dem Entscheidungsträger zusätzlich immer mehr disaggregierte Daten angeboten werden? Mit Blick auf die Publikation von Daten aus dem Rechnungswesen wird geprüft, welche Verbesserung eine Offenlegung von zusätzlichen Detailinformationen bei den Entscheidungen externer Bilanzadressaten bewirkt.³⁶⁾ Die Effizienzvermutung wird bei dieser Hypothesenvariante ähnlich begründet wie bei der Angebotsmenge: Den nur degressiv zunehmenden positiven Wirkungen der zusätzlichen, immer detaillierter werdenden Informationen stehen progressiv zunehmende negative Wirkungen gegenüber.

c) Aggregation als Komplexitätsreduktion:

Eine dritte, allgemeinere Hypothesenableitung unterstellt, daß die Entscheidung durch Informationsverdichtung zunehmend vereinfacht wird: Die Entscheidung weist einen abnehmenden Grad an **Komplexität** auf. Der umgekehrt

u-förmige Zusammenhang zwischen Aggregationsgrad und Entscheidungsqualität wird dann als ein **Spezialfall** des generell postulierten umgekehrt u-förmigen Zusammenhanges zwischen Komplexität und Effizienz angesehen.³⁷⁾ In ähnlicher Weise hätten wir bereits die Angebotsmenge als eine Operationalisierung der Umweltkomplexität ansehen können, wie dies namentlich von Schroder, Driver und Streufert getan wird.³⁸⁾

Kritisch ist zu allen drei Ableitungen anzumerken, daß sie die individuell verschiedenartig ausgeprägten Reaktionen auf unterschiedlich aggregierte Daten vernachlässigen. Eine Ausnahme bildet die Konzeption von Schroder, Driver und Streufert. Sie berücksichtigt unterschiedlich ausgeprägte u-förmige Verläufe in Abhängigkeit von dem Persönlichkeitsmerkmal "Kognitive Komplexität". Wir werden auf diese Differenzierung bei der Bestandsaufnahme der Forschung über "Kognitive Stile" und "Test-Intelligenz" eingehen.

Eine weitere Verfeinerung des Konzeptes ergibt sich aus der Tatsache, daß aggregierte und disaggregierte Informationen **hierarchisch geordnet** werden können. Der Entscheidungsträger kann sich aufgrund seiner Analyse höher aggregierter Daten auf bestimmte disaggregierte Informationen konzentrieren. Sein Differenzierungs- und (Re)-Integrationsprozeß kann insbesondere durch **baumstrukturartig aufgebaute Informationsangebote** unterstützt werden. Diese Gestaltungsmöglichkeit wird im nachfolgenden Gliederungsabschnitt über die Wirkung unterschiedlicher Präsentationsformen untersucht.

2.2.3.2. Die Ergebnisse: Gewisse Anzeichen für eine Bestätigung des umgekehrt u-förmigen Verlaufes

Unsere Bestandsaufnahme kann sich auf acht empirische Studien, die den Zusammenhang von Aggregation und Effizienz betriebswirtschaftlicher Entscheidungen untersuchen, stützen. Wie die tabellarische Bestandsaufnahme in Abbildung 2.5. zeigt, sind sowohl ihre Ergebnisse als auch ihre Aufgabenstellungen, Subjekte und Operationalisierungen von unabhängigen und abhängigen Variablen sehr unterschiedlich. Solche Unterschiede sollte eine Meta-Analyse grundsätzlich durchaus in Kauf nehmen, denn jeder Unterschied verkörpert eine spezifische Alternativhypothese, die man empirisch prüfen

kann.³⁹⁾ Gerade wegen der unterschiedlichen Operationalisierungen eines bestimmten theoretischen Konstruktes wird auf die Chance einer sog. "konzeptuelle Replikation"⁴⁰⁾ hingewiesen, die darin besteht, jeweils andere Facetten eines gemeinsamen theoretischen Konstruktes zu überprüfen.

Andererseits handelt es sich nicht um ein prinzipielles, sondern um ein graduell unterschiedlich abgestuftes Problem, das entsprechend abgestufte Maßnahmen erfordert. Im vorliegenden Fall ist es besonders schwierig, die Studien auf einer gemeinsamen Skala des untersuchten Aggregationsgrades einzuordnen. Wir weichen deshalb von unserer bisherigen Praxis ab, eine Gesamttendenz auszuzählen. Wir betrachten stattdessen zunächst die Studien, in denen drei und mehr Stufen des Aggregationsgrades variiert werden, denn nur bei ihnen kann man die umgekehrt u-förmige Hypothese testen. Danach wenden wir uns den Studien zu, die lediglich zwei unterschiedlich ausgeprägte Aggregationsgrade vergleichen.

a) Studien, die mindestens drei Stufen der Skala "Aggregationsgrad" vergleichen:

Drei der vier Studien, die mehr als drei Stufen der Skala "Aggregationsgrad" vergleichen, zeigen den postulierten umgekehrt u-förmigen Verlauf, jedoch zeigt sich nur bei einem Teil der vorgenommenen Vergleiche zwischen dem mittleren und den übrigen Aggregationsniveaus ein signifikanter Unterschied. Man kann daher nur von einer **tendenziellen Bestätigung** sprechen. Im einzelnen sind folgende Ergebnisse zu berichten:

1. Abdel-khalik untersucht, wie sich das Verhältnis der geschätzten Wahrscheinlichkeit der Einstellung von Zins- und Tilgungszahlungen ("default on loans") gegenüber der Nicht-Einstellung mit zunehmendem Angebot disaggregierter Daten verändert. Die von ihm ausgewiesenen t-Statistiken gelten jedoch nur **innerhalb** der Aggregationsniveaus. Rechnet man sie in punktbiseriale Korrelationskoeffizienten um,⁴¹⁾ so zeigt sich folgendes Ergebnis: Bei der Beurteilung der beiden Firmen, die ihre Zahlungen einstellten (was die Vpn aber nicht wußten), zeigt sich eine leicht zunehmende Tendenz mit zunehmender Menge an disaggregierten Informationen: Die Korrelationen steigen von 0,44 auf 0,60 bzw. von 0,68 auf 0,75. Bei den beiden Firmen,

Abbildung 2.5.: Einflüsse des Aggregationsgrades auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Barefield (Aggregation 1972)	Analyse von Kostenabweichungen, Klassifikation der Abweichungsursache	28 Stud.	Disaggregierte Kostenabweichungen vs. aggregierte Kostenabweichung; zweite exp. Bed.: Grad der Korrelation der Abweichungsursachen	Differenz zwischen tatsächlicher und lt. Modell optimaler Anzahl korrekter Klassifikationen	n.s. (mit disaggr. Information besser, Re-Analyse von "Table 2" punktbis. Korr.: 0,21)
Abdel-khalik (Aggregating Reports 1973)	Beurteilung, ob ein Firmenkunde Tilgungen und Zinszahlungen einstellt	207 Kreditmanager	drei exp. Bed.: bilanzielle Urdaten, teilw. verdichtete Daten, auf 12 Summenpositionen verdichtete Daten	Verhältnis der geschätzten Wahrscheinlichkeit von Zahlungseinstellg. und Nicht-Einstellung (vgl. "Table 25")	Geschätztes Verhältnis entwickelt sich unterschiedlich für beurteilte Firmen, die Zahlungen einstellen bzw. nicht einstellen (Re-Analyse s. Text)
Chervany/Dickson (Overload 1974)	Unternehmensspiel, Beschaffung und Produktion	22 Stud.	unverdichtete vs. verdichtete Berichte (nur Statistiken wie Mittelwert, Variationskoeffizient), 1040 vs. 446 "data-items"	Kosten	Sign. geringere Kosten bei verdichteten Berichten, keine Effektschätzung wg. mangelnder Inf.
Senn (Information System 1973)	Unternehmensspiel, Beschaffung	28 Einkäufer	EDV-Liste aggregiert vs. EDV-Liste detailliert (dritte exp. Bed.: verdichtete Bildschirminformation)	Kosten ("Total Costs")	n.s., (Re-Analyse von "Table 13", erklärte Varianz: 3,8 v. H.)
Smith (Segment Reporting 1979)	Prognose des Betriebserfolges	Finanzanalysten, Anzahl unbek.	ohne Zusatzinformationen vs. wenige aggregierte Segmentinformationen vs. zahlreiche detaillierte Segmentinformationen	Absolutbetrag des prozentualen Schätzfehlers	beste Prognose bei aggreg. Segmentinformation, sign. Unterschiede zw. erster u. zweiter exp. Bed.
Casey (Information Load 1980)	Beurteilung, ob ein Unternehmen insolvent wird innerhalb der nächsten drei Jahre	122 Kreditmanager	sechs ausgewählte Kennzahlen vs. zusätzlich Bilanz und Erfolgsrechnung vs. zusätzliche Erläuterungen	Anzahl der korrekt klassifizierten Firmen	Beste Klassifikation bei mittlerem Aggregationsniveau, Diff. zw. Kennzahlen und Bilanzdaten sign., zw. Bilanzdaten und Erläuterung nicht (Re-Analyse von "Table 4", erklärte Varianz: 24,4 v. H.)
Davis (Information Presentation 1980)	Unternehmensspiel, Abstimmung von Produktions- und Absatzmengen	96 Stud.	Rohdaten vs. verdichtete Daten (Mittelwerte, Standardabweichung etc.), weitere exp. Bed.: Tabelle vs. Graphik	Kosten	Rohdaten sign. besser als verdichtete Daten (Re-Analyse des F-Wertes Text S. 54 unten: punkt-biseriale Korrelation 0,36), kein Interaktionstest
Otley/Dias (Aggregation 1982)	Unternehmensspiel, Produktprogrammplanung	40 Stud.	vier verschiedene Aggregationsgrade der Berichte, u-förmige Beziehung zum Schätzfehler behauptet	Schätzfehler von 1. Kostenprognose ("DIEE") 2. Gewinnprognose ("CPEE")	1. sign. u-förmige Beziehung, erkl. Varianz 30 v. H., 2 von 4 Kontrasten sign. (eins. Test) 2. u-förmige Beziehung, tend. best. erkl. Varianz 12 v. H., 3 von 4 Kontrasten sign. (eins. Test)

die ihre Zahlungen nicht einstellen, zeigt sich die umgekehrte Tendenz: Die Korrelationen nehmen von 0,43 auf 0,40 bzw. von 0,25 auf 0,20 ab. Eine eindeutige Schlußfolgerung läßt sich hieraus nicht ableiten. Wegen der uneinheitlichen Entwicklung der Korrelationskoeffizienten haben wir von einem Signifikanztest auf Unterschiedlichkeit der Korrelationen abgesehen. Das recht grobe "Effizienzmaß" der Fragebogen-Studie von Abdel-khalik ist unseres Erachtens auch nicht frei von subjektiven Risikopräferenzen. Seine Studie wird daher nur als bedingt geeignet angesehen, die uns interessierende Effizienzhypothese zu prüfen.⁴²⁾

2. Smith untersucht, welchen Einfluß nach Produktgruppen "segmentierte" Bilanzen auf die Prognose des Betriebserfolges divisionalisierter Unternehmen haben. Ihre Studie zeigt, daß sich mit wenigen aggregierten Zusatzinformationen die Prognose signifikant verbessert. Wenn disaggregierte Produktgruppen-Informationen hinzukommen, ergibt sich zwar die erwartete Verschlechterung der Prognose, aber dieser Effekt ist nicht signifikant.

3. Casey läßt Kreditmanager klassifizieren, ob ein Unternehmen innerhalb der nächsten drei Jahre insolvent wird oder nicht. Seine Untersuchung zeigt einen signifikanten Anstieg der Klassifikationsgüte, wenn außer den sechs Kennzahlen die numerischen Bilanzdaten angeboten werden. Ein zusätzliches Angebot von verbalen Erläuterungen ("footnotes") zeigt einen schwachen, nicht signifikanten Abfall.

4. Otley und Dias nehmen eine entscheidungstheoretisch abgesicherte Variation des Aggregationsgrades vor. Beim Übergang von der ersten zur zweiten und von der zweiten zur dritten Stufe bleibt die für eine Produktprogrammplanung entscheidungsrelevante Information zur Bestimmung der optimalen Lösung (mithilfe eines LP's) erhalten, der Umfang der Berichte wird jedoch von neun auf fünf Seiten eingeschränkt. Erst beim Übergang von der dritten auf die vierte Stufe findet eine Zusammenfassung von fixen und variablen Kostenbestandteilen statt, die den entscheidungstheoretischen Wert der Informationen beeinträchtigt. Die Schätzfehler von Kosten- und Gewinnwerten zeigen einen u-förmigen Verlauf. Da es sich um ein Fehlermaß handelt, entspricht dies dem behaupteten umgekehrt u-förmigen Verlauf der Prognosegüte. Die erklärte Varianz beider Verläufe ist mit durchschnittlich 21 v. H. als

beachtlich anzusehen. Die Mehrzahl der getesteten Kontraste ist signifikant ausgeprägt. Daß nicht alle Kontraste signifikant sind, sollte angesichts der geringen Stichprobengröße noch nicht als Falsifikation der wissenschaftlichen Hypothese angesehen werden.

Insgesamt gesehen liefern die drei zuletzt referierten Studien eine tendenzielle Bestätigung der umgekehrt u-förmigen Hypothese. Vergleichsweise hoch ist die Validität der Operationalisierung der unabhängigen Variablen in der Arbeit von Otley und Dias. Die beiden anderen Untersuchungen können eigentlich nur den aufsteigenden Ast der Kurve testen. Um zu prüfen, ob sich bei wachsendem Angebot von zunehmend stärker disaggregierten Informationen tatsächlich dysfunktionale "Über"-Lastungseffekte zeigen, hätte man die **Nutzung** der angebotenen Informationen prozeßbegleitend erheben und bei der Auswertung berücksichtigen müssen.⁴³⁾

b) Studien, die nur zwei Stufen der Skala "Aggregationsgrad" vergleichen:

Die vier weiteren Studien zeigen konfliktäre Ergebnisse:

1. Barefield stellt bei einem einfachen Beurteilungsproblem keine signifikanten Unterschiede fest. Wegen der recht geringen Varianzerklärung⁴⁴⁾ von ca. 7 v. H. ist der Befund trotz des geringen Stichprobenumfanges als eine Falsifikation der Hypothese zu werten.

2. Senn stellt bei einem Unternehmensspiel aus dem Beschaffungsbereich ebenfalls keine signifikanten Unterschiede fest. Auch seine Studie ist wegen der geringen Varianzerklärung von 3,8 v. H. als Falsifikation anzusehen.

3./4. Richtungskonträr sind die signifikanten Befunde von Chervany/Dickson und Davis ausgeprägt. Die angebotenen Berichte in ihren Unternehmensspielen enthalten entweder unverdichtete Rohdaten oder zusammenfassende Statistiken (Mittelwert, Streubereich etc.). Da bei Chervany/Dickson in der Rohdaten-Bedingung sehr viele Informationen angeboten werden, ist zu vermuten, daß der absteigende Ast der Kurve geprüft wird. Es fehlen uns jedoch numerische Angaben, die belegen, daß bei Davis der ansteigende Ast der Kurve vorliegt.

2.2.3.3. Zusammenfassende Würdigung

Faßt man die Ergebnisse der referierten Studien zusammen, so läßt sich die umgekehrt u-förmige Hypothese, zumindest, soweit sie sich auf die erste Ableitungsvariante stützt, nicht verwerfen. Die empirische Evidenz ist jedoch weder einheitlich noch außerordentlich stark ausgeprägt. Es fehlen auch bei diesem Gestaltungsparameter differenzierende Untersuchungen, die klären, wie der Aggregationsgrad das Informationsverhalten beeinflußt und warum sich unterschiedliche Effizienzwerte ergeben (sollen). Es ist somit auch hier ein Defizit an verhaltensbezogener Forschung festzustellen.

Die divergierenden Ergebnisse erlauben es auch bei diesem Gestaltungsparameter nicht, Handlungsanweisungen auszusprechen. Die Betreiber von Managementinformationssystemen sollten jedoch im Einzelfall sorgfältig prüfen, ob sie nicht ohne Hinnahme eines Verlustes an Entscheidungsqualität eine **Verbesserung der Prozeßeffizienz** erreichen können. Daß die Entscheidungszeit mit zunehmender Menge an angebotenen disaggregierten Daten signifikant steigt, wird nämlich durch eine Reihe von Studien belegt.⁴⁵⁾ Durch eine Verringerung des Angebotes könnten auch Kosten bei der Erfassung und Speicherung von Daten eingespart werden.

2.2.4. Angebotsmedium und Effizienz

Die empirischen Studien zum Einfluß des Angebotsmediums richten sich auf zwei Fragen:

1. Bewirkt eine Präsentation der Daten auf Bildschirmen eine andere Effizienz als ein Datenangebot auf einer "hard-copy" (EDV-Liste)?
2. Ist eine individuelle Abfrage von Informationen aus einer Datenbank, die sich nach den Informationsbedürfnissen eines Benutzers richtet, einem vorstrukturierten, fest-formatierten Informationsangebot in tabellarischer Darstellung überlegen?

Die theoretische Fundierung dieser Effizienzvermutungen ist dürftig.⁴⁶⁾ Die durchgeführten Studien haben eher explorativen Charakter. Sie dürfen jedoch für sich in Anspruch nehmen, daß sie die Wirksamkeit dieser Gestaltungsparameter überhaupt einer empirischen Prüfung unterzogen haben.

Das Ergebnis der wenigen durchgeführten Effizienzprüfungen ist ernüchternd (vgl. Abbildung 2.6): Fünf der sechs Studien zeigen keine signifikanten Unterschiede, in der sechsten, von Lucas durchgeführten Untersuchung, erweist sich das traditionelle Medium EDV-Liste signifikant überlegen. Auch in der nachrichtlich ausgewiesenen Untersuchung von Weigand, die kein ökonomisches Effizienzmaß benutzt, schneidet die traditionelle Präsentationsform besser ab.⁴⁷⁾

Man muß bei diesem Ergebnis berücksichtigen, daß die Stichprobenumfänge der Untersuchungen mit nicht-signifikanten Ergebnissen sehr klein sind. Es hätte schon starker Effekte in den zugehörigen Populationen bedurft, um die statistischen Nullhypothesen zu verwerfen. Trotzdem erteilen die Ergebnisse einer "un"-bedingten Fortschrittsgläubigkeit eine klare Absage: Die Lage der Mittelwerte ist meistens so ausgeprägt, daß die konventionelle Lösung etwas besser abschneidet. Bei Smith konnten wir eine Effektschätzung aufgrund der ausgewiesenen Rohdaten vornehmen: Die punkt-biseriale Korrelation beträgt nur 0,03. Dies entspricht einer erklärten Varianz von weniger als 1 v. H. Man kann also nicht davon ausgehen, daß es nur eines größeren Stichprobenumfanges bedurft hätte, um die Überlegenheit der neuen Technologien nachzuweisen. Ergänzungsbefunde zum Verlauf der Effizienzunterschiede über mehrere Perioden der durchgeführten Unternehmensspiele deuten auch nicht daraufhin, daß sich nach einer relativ kurzen Lernphase bereits eine Überlegenheit der neuen Medien herausstellt.⁴⁸⁾

Man kann auch nicht davon ausgehen, daß die Bildschirmarbeit den Entscheidungsprozeß signifikant beschleunigt: Kozar ermittelt einen signifikant geringeren Zeitverbrauch für die "hard-copy", Senn und Jenkins stellen keine signifikanten Unterschiede fest.⁴⁹⁾ Die Möglichkeit der (zusätzlichen) Nutzung einer individuell abfragbaren Datenbank bewirkt bei Benbasat und Dexter eine signifikante Verlängerung des Informationsprozesses, Smith

Abbildung 2.6.: Einflüsse des Angebotsmediums auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Kozar (Decision Making 1972)	Unternehmensspiel, Beschaffung und Produktion	22 Stud.	EDV-Liste aggregiert vs. Bildschirmausgabe aggregiert	Kosten ("Total Cost")	n.s., keine Effektschätzung wg. mangelnder Informationen
Senn (Information System 1973)	Unternehmensspiel, Beschaffung	28 Einkäufer	EDV-Liste aggregiert vs. Bildschirm aggregiert (dritte exp. Bed.: EDV-Liste detailliert)	Kosten ("Total Costs")	n.s., keine Effektschätzung wg. mangelnder Informationen
Smith (Inquiry Techniques 1975)	Unternehmensspiel, Beschaffung und Produktion	17 Stud.	Bildschirmausgabe mit vorgegebenen Berichten vs. zusätzlich indiv. Berichte mittels Datenbanksystem (tabell. vs. graphische indiv. Berichte)	1. Kosten ("Total Cost") 2. Prognosegenauigkeit	1. n.s., punkt-bis. Korrr.: 0,03 (Re-Analyse Tab. 7.1) 2. n.s., keine Effektschätzung wg. mangelnder Inf.
Jenkins (Design Variables 1977)	Unternehmensspiel, Produktion	32 Stud.	EDV-Liste vs. Bildschirmausgabe	Kosten ("average cost")	n.s., keine Effektschätzung wg. mangelnder Inf.
Benbasat/Dexter (Value and Events 1979)	Unternehmensspiel, Beschaffung und Produktion	24 Stud. 4 Manager 20 Fak. Mitgl.	Strukturierter, aggregierter Bericht vs. individuelle Abfragen im Dialog	Deckungsbeiträge	n.s., keine Effektschätzung wg. mangelnder Informationen
Lucas (Graphics 1981)	Unternehmensspiel, Beschaffung	44 Manager	EDV-Liste vs. Bildschirm, beide tabellarisch	Kosten ("best simulation")	EDV-Liste sign. besser, (Re-Analyse von "Table 1", punkt-bis. Korrr.: 0,26)
Weigand (Mensch-Maschine-Kommunikation 1983)*	Eignungstest zur Personalbeurteilung	33 Stud.	"Papier-und-Bleistift"-Version vs. Bildschirm-Version ("Online-Form")	Punktezahl ("Gesamtergebnis")	Papier-Version sign. besser (eins. Test), (Re-Analyse: punkt-bis. Korrr.: 0,30)

*: Studie wird nur nachrichtlich vermerkt, da kein ökonomisches Effizienzmaß

beobachtet keine signifikanten Unterschiede.⁵⁰⁾ Ergänzungsbefunde der referierten Studien belegen, daß die Entscheidungsdauern pro Spielperiode im Zeitablauf stark abnehmen. Die Lernfortschritte spiegeln jedoch nur die wachsende Vertrautheit der Versuchsteilnehmer mit dem Problem wider: Sie verlaufen bei beiden Medien praktisch parallel.⁵¹⁾

Sicherlich kann man aus den durchgeführten Studien nicht schließen, daß das "Medium" der Informationspräsentation für die Effizienz von komplexen Entscheidungs- und Beurteilungsprozessen unbeachtlich sei. Die bisher durchgeführten Studien stützen sich teilweise auf eine EDV-Technologie, die nach heutigen Maßstäben als veraltet anzusehen ist. Der Vergleich von "Bildschirm" vs. "Liste" findet zudem in einer historischen Phase statt, in der die meisten Benutzer noch keine intensiven Erfahrungen in der Mensch-Maschine-Kommunikation gesammelt hatten. Deshalb halten wir es auch nicht für zweckmäßig, aus den referierten Studien praxeologische Gestaltungsempfehlungen abzuleiten.

Andererseits macht man es sich unseres Erachtens zu leicht, wenn man wegen des Hinweises auf "veraltete Instrumente" aus der bisherigen Forschung keine Lehren für zukünftige Effizienzprüfungen zieht. Die nicht-signifikanten Ergebnisse der explorativen Studien belegen recht deutlich, daß man den Vergleich von unterschiedlichen Problemlösungen nicht so global vornehmen darf. Wer heutzutage mit wissenschaftlichem Anspruch behauptet, daß eine neuartige Problemlösung einer konventionellen überlegen sei, der muß dies auch theoretisch begründen und die Voraussetzungen nennen, unter denen seine Behauptungen gelten sollen. Nur so wird eine angemessene empirische Prüfung und ein wissenschaftlicher Fortschritt möglich. Gerade weil es im Zeitalter des "personal" Computer durchaus neuartige Hard- und Software-Entwicklungen gibt, ist eine differenzierte Erforschung geboten. Die Differenzierung sollte sich nicht nur auf die gleichzeitige Variation mehrerer Parameter des Informationssystems erstrecken - wie dies bereits in der anspruchsvollen Pionierstudie von Benbasat geschieht. Sie sollte auch das menschliche Informationsverhalten differenziert berücksichtigen.

2.2.5. Graphische Unterstützung und Effizienz

2.2.5.1. Die Hypothesen

Der Einfluß graphischer Unterstützungshilfen des Informationsangebotes wird, wie der ausgezeichnete Überblicksartikel von DeSanctis⁵²⁾ belegt, schon seit längerer Zeit erforscht. Auswirkungen auf die Qualität betrieblicher Entscheidungen wurden jedoch erst in jüngerer Vergangenheit untersucht, angeregt durch die gesteigerten graphischen Möglichkeiten von Computern. Da mittlerweile auch gängige "personal computer" über gute Graphikhard- und -software verfügen, kommt graphischen Hilfen eine besondere Bedeutung zu. Die empirischen Prüfungen der uns interessierenden Effizienzgrößen richten sich vornehmlich auf den Vergleich folgender Instrumente:⁵³⁾

- a) "Tabellen vs. Graphiken",
- b) "Tabellen vs. Schematische Gesichter",
- c) "Tabellen vs. Bäume" und
- d) "einfarbige vs. mehrfarbige Präsentation".

a) "Tabellen vs. Graphiken":

Unter dieser Rubrik sind die meisten bisher durchgeführten Effizienzprüfungen einzuordnen. Sie vergleichen die Wirkungen von numerischen Daten, die in Tabellen ausgewiesen werden, mit graphischen Darstellungen dieser Daten. Bis auf Zmud, Blocher und Moffie, die farbige Balkendiagramme verwenden,⁵⁴⁾ werden **Kurvendarstellungen** benutzt. Sie veranschaulichen Beziehungen zwischen Variablen oder Entwicklungen im Zeitablauf. Während man in älteren Untersuchungen noch recht grobe "Treppen"-Graphiken⁵⁵⁾ einsetzt, um Kurvenverläufe darzustellen, können die neueren Untersuchungen auf kommerziell verfügbare Computergraphiken zurückgreifen, in denen kontinuierliche Kurvenzüge auf hochauflösenden, großformatigen Displays in mehreren Farben dargestellt werden.⁵⁶⁾ Vor allem bei diesen modernen Informationstechnologien wird die Hypothese vertreten, daß Graphiken Entscheidungsprozesse beschleunigen und die Qualität der Entscheidung verbessern, indem sie den Blick des Benutzers auf die wesentliche Information lenken.⁵⁷⁾

b) "Tabellen vs. Schematische Gesichter":

Die auf Chernoff zurückgehenden "schematic faces" werden eingesetzt, um multivariate Eigenschaftsvektoren zu einem leicht und prägnant wahrnehmbaren Gesamturteil zusammenzufassen.⁵⁸⁾ In der verhaltensorientierten Rechnungswesenforschung wird geprüft, ob dieses Hilfsmittel geeignet ist, die Güte und Geschwindigkeit der klassifikatorischen Prognose charakteristischer Unternehmensentwicklungen zu verbessern.

c) "Tabellen vs. Baumstrukturen":

Baumstrukturen unterscheiden sich von den bisher dargestellten graphischen Hilfsmitteln dadurch, daß die numerischen Informationen weiterhin als Zahlen dargestellt werden. Geändert wird jedoch ihre Anordnung auf einem Informationsträger, so daß die hierarchischen **Beziehungen** zwischen den Zahlen hervorgehoben werden.⁵⁹⁾ In der betriebswirtschaftlichen Praxis und in der Fachliteratur hat es sich schon seit langem eingebürgert, Kennzahlensysteme in Form von Baumstrukturen darzustellen. Unseres Wissens ist jedoch erst von Hauschildt systematisch empirisch geprüft worden, ob eine derartige Präsentationsform geeignet ist, die Effizienz eines komplexen Beurteilungsprozesses zu erhöhen.⁶⁰⁾ Seine Effizienzvermutung stützt sich auf die Überlegung, daß die Baumstruktur in besonderer Weise geeignet ist, die bei einem komplexen Beurteilungsproblem vorzunehmenden Differenzierungen und Integrationen zu unterstützen.

b) "Einfarbige vs. mehrfarbige Präsentation":

Sowohl Tabellen als auch Graphiken können mono- oder multichromatisch präsentiert werden. Darüber hinaus kann man unterscheiden, ob die Farben für eine redundante oder nicht-redundante Codierung eingesetzt werden. Im zweiten Fall werden durch andere optische Hilfsmittel bereits abgegrenzte Merkmale zusätzlich durch unterschiedliche Farben voneinander abgehoben. In den hier referierten neueren Studien von Benbasat und seinen akademischen Kollegen wird die zweite Möglichkeit gewählt. Für sie soll in besonderem Maße gelten, daß sie die Lesbarkeit von Darstellungen erhöht und damit geeignet ist, das Erkennen problemrelevanter Informationen zu unterstützen.⁶¹⁾

Zusammengefaßt werden folgende Hypothesen geprüft:

1. Eine Darstellung in:

- a) Kurvenzügen oder Balkendiagrammen,
- b) Schematischen Gesichtern,
- c) Baumstrukturen,

führt zu einer größeren Effizienz als eine Darstellung in Tabellen.

2. Eine mehrfarbige Darstellung führt zu einer höheren Effizienz als eine einfarbige Darstellung.

2.2.5.2. "Tabelle vs. Graphik": Keine praktisch bedeutsamen Unterschiede

Das empirische Ergebnis steht in krassem Gegensatz zu den hochgesteckten Erwartungen, die eine aggressive Werbung für kaufmännische Computergraphik bei potentiellen Benutzern weckt: **Keine der 12 referierten Studien zeigt eine signifikante Überlegenheit graphischer Darstellungen (vgl. Abbildung 2.7).** Die von uns nachträglich ermittelten Varianzerklärungen liegen unter 10 v. H.

Nur in der Studie von Vent ergibt sich ein tendenziell signifikanter Haupteffekt mit einer Varianzerklärung von 13,8 v. H. Dieser Effekt bezieht sich jedoch nicht auf einen Vergleich von "Tabelle vs. Kurvendarstellung", sondern auf zwei unterschiedliche Einführungsveranstaltungen, bei denen in kurzer Zeit zahlreiche Tabellen bzw. zahlreiche Graphiken angeboten werden ("Tabelle vs. Graphik bei Systemdarstellung"). Bezüglich der für das Feedback eingesetzten Computergraphiken bzw. -tabellen ergibt sich auch bei Vent kein signifikanter Effekt. Dem Befund von Vent steht ein Ergebnis von Remus gegenüber, der für bereinigte Kostenwerte eine signifikante Überlegenheit der Tabelle feststellt. Auch in der nachrichtlich vermerkten Studie von Lusk schneiden Tabellen signifikant besser ab. Schließlich ist auf die Dissertation von Ghani hinzuweisen, die nach DeSanctis und Ives ebenfalls eine Überlegenheit der Tabelle belegt.⁶²⁾

Abbildung 2.7.: Einflüsse graphischer Hilfsmittel auf die Effizienz betrieblicher Entscheidungen (Teil 1)

- Graphiken vs. Tabellen -

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Benbasat (Effects 1974)	Unternehmensspiel, Beschaffung und Produktion	32 Stud.	Tabellen vs. Graphiken (grobe "Treppen"-Kurven) andere exp. Bed.: Entscheidungshilfen, Angebotsmenge, "exception reporting"	Kosten	n.s., (Re-Analyse von Tab. 1 Korrr.: -0,28, erkl. Var. 7,6 v. H.; tend. sign. Interakt. mit Entscheidungshilfe)
Smith (Inquiry Techniques 1975)	Unternehmensspiel, Beschaffung und Produktion	17 Stud.	Erzeugen eigener (zusätzlicher Bildschirm Ausgaben) in tabellarischer oder graphischer Darstellung (grobe "Treppen"-Kurven)	1. Kosten 2. Prognosegenauigkeit	1. n.s., (Re-Analyse S. 119: punkt-bis. Korrr.: 0,04 2. n.s., keine Effektschätzg. wg. mangelnder Inf.
Lusk (Differential Peaking 1979)*	Beantwortung von Faktenfragen	300 Stud.	Zwei tabellarische und drei graphische Berichtsvarianten; Kognitiver Stil: GEFT	Anzahl korrekt beantworteter Faktenfragen	Tabelle sign. besser, keine sign. Interaktion mit Kogn. Stil
Davis (Information Presentation 1980)	Unternehmensspiel, Abstimmung von Produktions- und Absatzmengen	96 Stud.	Tabelle vs. Graphik, zweite exp. Bed. Rohdaten vs. verdichtete Daten (Mittelwerte, Standardabw. etc.)	Kosten	n.s., (keine Effektschätzung wg. mangelnder Informationen)
Lucas/Nielsen (Presentation 1980)	Unternehmensspiel, Logistik eines Handelsunternehmens	36 Stud. 78 Manager	tabellarische vs. graphische Darstellung	erzielte Gewinne ("mean profit")	n.s., (keine Effektschätzung wg. mangelnder Informationen)
Lucas (Graphics 1981)	Unternehmensspiel, Bestellmengenentscheidung unter Unsicherheit	97 Manager	tabellarische vs. nur graphische Bildschirm Ausgabe oder tabellarische und graphische Ausgabe ("Table 4"), Kognitiver Stil: Analytisch vs. Heuristisch (lt. Barkin)	Kosten ("best simulation")	n.s., (keine Effektschätzung wg. mangelnder Informationen; tend. sign. Interaktion mit Kognitivem Stil)
Ghani/Lusk (Information Representation 1982)	einfaches Unternehmensspiel, Absatzmengenentscheidung unter Unsicherheit	49 Stud.	Vier Gruppen: 1. Zuerst Tabelle, dann Graphik; 2. Zuerst Graphik, dann Tabelle; 3. Nur Graphik; 4. Nur Tabelle (alle Prés. farbig)	1. Gewinne 2. Gewinnveränderung	1. n. s. 2. n. s. Keine Effektschätzungen wg. mangelnder Informationen
Zmud/Blocher/Moffie (Color Graphic 1983)	Bewertung von Risiken bei Rechnungsprüfungen	51 interne Revisoren	Farbige Balkendiagramme vs. schwarz-weiße tabellarische Darstellung; Zweite experimentelle Bedingung: 5 vs. 9 Informationen, je 15 Sekunden Darbietungszeit	Ob die Risikoklasse der Rechnungen richtig eingestuft wird	n.s., (Re-Analyse von "Table 3", erkl. Var. 0,01 v. H., sign. Interaktion bei 5 Inf. Graphik besser bei 9 Informationen Tabelle)

Abbildung 2.7.: Einflüsse graphischer Hilfsmittel auf die Effizienz betrieblicher Entscheidungen (Teil 1)

- Graphiken vs. Tabellen (Fortsetzung) -

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Remus (Presentations 1984)	Unternehmensspiel, Abstimmung von Produktions- und Absatzmengen	53 Stud.	tabellarische vs. graphische Darstellung	Kosten 1. Tatsächliche Kosten ("actual costs") 2. Kosten auf Basis indiv. Regressionen der Entschdg. 3. Kosten auf Basis gemittelter Regressionsparameter	1. n.s., Korr.: -0,02 2. n.s., Korr.: -0,08 3. sign. geringere Kosten bei tabellarischer Darstellung, Korr.: -0,62 (Re-Analyse: punkt-bi-seriale Korr., ermittelt aufgrund mittlerer t-Werte der Tab. 1,2 und 3)
Benbasat/Dexter (Graphical Information Presentation 1985)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verschiedene Verkaufsgebiete	35 Stud.	tabellarische vs. graphische Darstellung (außerdem beide Präsentationsformen ohne vs. mit Farbe); außerdem Kognitiver Stil (GEFT)	Erzielte Gewinne, relativ zum Optimum	n.s., (keine Effektschätzung wg. mangelnder Informationen, keine sign. Interaktionseffekte)
Benbasat/Dexter/Todd (Graphical Information Presentation 1985)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verschiedene Verkaufsgebiete	66 Stud.	tabellarische vs. graphische Darstellung (außerdem beide Präsentationsformen ohne vs. mit Farbe)	Erzielte Gewinne, relativ zum Optimum	n.s., (keine Effektschätzung wg. mangelnder Informationen, keine sign. Interaktionseffekt)
Benbasat/Dexter (Time Constraints 1985)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verschiedene Verkaufsgebiete	65 Stud.	Tabelle vs. Graphik vs. Tabelle und Graphik (2. exp. Bed.: ohne vs. mit Farbe, 3. exp. Bed. Zeitlimit 5 vs. 15. Minuten, 3x2x2-Design)	Erzielte Gewinne, relativ zum Optimum, ob 99 v. H. erreicht werden	n.s. (Re-Analyse von Table 2 a) und b) zusammengefaßt, Chiquadrat-test, tetrachor. Korr.: -0,20; bei 15-Minuten-Bedingung Graphik sign. schlechter, bei 5 Minuten-Bedingung n.s.)
Vent (Denkstrategien 1985)	Energieversorgung der BRD, komplexes System, 150 Aktionsparameter können beeinflußt werden	24 Stud.	Tabelle vs. Graphik: a. bei Systemdarstellung b. bei Feedback (2x2-Design)	Gesamtindex über 17 zentrale Variablen, Kosten mit 50 v. H. gewichtet	a. Graphik besser, tend. sign., erkl. Varianz 13,8 v. H. b. n.s., erkl. Var. 7,8 v. H. tend. sign. Interakt. erkl. Var. 12,9 v. H. (Re-Analyse v. Tab. 2)

*: Studie wird nur nachrichtlich vermerkt, da kein ökonomisches Effizienzmaß

Es ist daher Lucas, Ives, DeSanctis, Benbasat und Ko-Autoren zuzustimmen, daß man nicht generell davon ausgehen darf, daß Graphiken Tabellen überlegen seien. Auch die Umkehrung gilt nicht. Vielmehr ist der von diesen Autoren geforderte hypothesengeleitete **Kontingenzansatz** bezüglich Aufgaben-, Personen- und System-Merkmalen zu verwirklichen. Theoretische Überlegungen entwickelt DeSanctis. Erste Ergebnisse eines solchen Forschungsprogrammes legen Benbasat und Kollegen in den von uns referierten Arbeiten vor. Die von ihnen und anderen bestimmten **Interaktionseffekte** sind jedoch noch nicht hinreichend belegt, um hierauf bereits Gestaltungsempfehlungen aufzubauen. So kann man aus Abbildung 2.7. zwar folgende signifikante oder tendenziell signifikante Interaktionen ablesen:

- * mit Entscheidungshilfe steigert eine Graphik die Effizienz (Benbasat, Effects 1974);
- * "Heuristiker" werden durch Graphiken besonders gefördert (Lucas, Graphics 1981);
- * bei einfachen Aufgaben bewirken Balken-Graphiken höhere Klassifikationsgüten, bei komplexen Problemstellen liefern Tabellen bessere Ergebnisse (Zmud/Blocher/Moffie, Color Graphic 1983);
- * bei fehlendem Zeitdruck weisen Tabellen günstigere Ergebnisse auf (Benbasat/Dexter, Time Constraints 1985);
- * ein Wechsel der Präsentationsform bewirkt eine Verschlechterung (Vent, Denkstrategien 1985).

Aber: Diesen Befunden stehen auch **nicht-signifikante** Ergebnisse gegenüber:

- * keine signifikanten Interaktionen bei drei (anderen) Systemvariablen und zwei Persönlichkeitsvariablen (Benbasat, Effects 1974),
- * keine signifikante Interaktion mit dem Wechsel der Präsentationsform (Ghani/Lusk, Information Representation 1982).

* keine signifikanten Interaktionen mit dem Kognitiven Stil (Benbasat/Dexter, Graphical Information Presentation 1985),

* keine signifikanten Interaktionen mit dem Faktor "ein- vs. mehrfarbig" (Benbasat/Dexter, Graphical Information Presentation 1985, Benbasat/Dexter/Todd, Graphical Information Presentation 1985).

Solange nicht geklärt ist, auf welchen intervenierenden, bisher nicht beobachteten Prozessen diese divergierenden Ergebnisse beruhen, scheint es uns verfrüht, praxeologische Schlußfolgerungen aus den beobachteten Interaktionen zu ziehen.

Gegen unsere Bestandsaufnahme läßt sich einwenden, daß Graphiken auch dann noch die Wirtschaftlichkeit der Entscheidung steigern, wenn sie bei gleicher Entschlußqualität eine Verringerung des Entscheidungsaufwandes bewirken. Dies wird in experimentellen Studien an dem leicht und zuverlässig erhebbaren **Zeitverbrauch** für die Entscheidung gemessen. Unsere diesbezügliche Zusatzauswertung läßt **keine Überlegenheit der Graphik** erkennen: In sechs Studien ergeben sich keine signifikanten Unterschiede⁶³⁾. Nur in der Untersuchung von Benbasat und Dexter, in der eine Zeitbegrenzung vorgegeben wird, zeigt sich ein signifikant geringerer Zeitverbrauch in der Graphik-Gruppe.⁶⁴⁾ Nach DeSanctis soll sich bei der Studie von Ghani ein geringerer Zeitverbrauch bei der Tabelle ergeben haben.⁶⁵⁾ Somit ist auch bei diesem Merkmal die Bilanz ausgeglichen.

Offen bleibt die Frage, **warum** sich Tabellen und Graphiken bezüglich der Effizienzwirkungen so selten signifikant unterscheiden. Unseres Erachtens kann dies zwei Ursachen haben:

1. Die Präsentationsform beeinflusst lediglich **Wahrnehmungsprozesse**, deren Wirkungen für die Entscheidungsqualität oder den gesamten Zeitverbrauch der Entscheidung vergleichsweise unbeachtlich sind.

2. Die Entscheidungsträger verfügen über praktisch **äquifunktionale Aufnahme- und Verarbeitungsmechanismen**, um die in Tabellen und Graphiken enthaltene Information zu dekodieren.

Zu 1.: In den bisherigen Bestandsaufnahmen wird die erste Erklärung favorisiert. Insbesondere wird darauf verwiesen, daß es von der **Aufgabe** abhängt, welche Funktionen eine Graphik oder Tabelle unterstützen könne. Komplexe Entscheidungsprobleme stellen jedoch regelmäßig multi-funktionale Anforderungen. Will man sich nicht auf simple Teil-Aufgaben zurückziehen, dann kommt man leicht in Bewertungskonflikte, wenn man eine Prognose über die relative Wirksamkeit von Graphiken oder Tabellen abgeben will, weil bestimmte Funktionen von Tabellen, andere von Graphiken besser erfüllt werden.⁶⁶⁾

Zu 2.: Die Studie von Ghani und Lusk belegt, daß dieselben Entscheidungsträger nach einer gewissen Umstellungsphase durchaus in der Lage sind, ein bestimmtes Entscheidungsproblem mit einer alternativen Präsentationsform der angebotenen Daten gleich gut zu lösen.⁶⁷⁾ Dies gilt für eine Umstellung von Tabellen auf Graphiken und Graphiken auf Tabellen in gleicher Weise. Wir sind deshalb nicht der Meinung, daß eine Überlegenheit der Graphik nur eine Frage einer angemessenen Einübungszeit und eines geschickten Marketing für ihre Akzeptanz sei.⁶⁸⁾

Wir sind überdies der Auffassung, daß man, teilweise beziehungsweise auf populärwissenschaftliche Darstellungen, die relative Hirnhälften-Dominanz unterschiedlicher "Typen" von Menschen zur Erklärung der differentiellen Wirkung von Graphiken und Tabellen zu sehr betont hat. Beide Hirnhälften **interagieren** intensiv miteinander, beide sind an der Entschlüsselung numerischer und analoger Reize beteiligt. Daher ist es auch unterschiedlichen Typen von Menschen möglich, beide Arten von Informationen innerhalb der hier relevanten Grenzen äquifunktional zu nutzen. Dies ist unseres Erachtens der Grund, warum man bisher keine konsistenten Wechselwirkungen von Präsentationsform, Kognitivem Stil und Effizienz feststellen konnte.⁶⁹⁾

Es scheint uns deshalb auch bei der Analyse von "Graphiken vs. Tabellen" zweckmäßig, prozeßbegleitende Meßverfahren einzusetzen, die differenziert untersuchen, **wie** Graphiken und Tabellen das Informationsverhalten beeinflussen. Darüber hinaus sollte man den Entscheidungsträgern die Möglichkeit geben, die Informationspräsentation nach ihren eigenen Informationsbedürfnissen selbst zu **gestalten**, denn dies entspricht der realen Situation.⁷⁰⁾ Der besondere Vorteil moderner Computergraphiken liegt unseres Erachtens gerade in den vielfältigen Kombinationsmöglichkeiten, die es auch graphischen Laien ermöglichen, vergleichsweise anspruchsvolle graphische oder tabellarische Darstellungen zu erzeugen. Um dies zu erforschen, muß man allerdings das sich dynamisch entfaltende, prozeßbegleitend gemessene **Informations-Verhalten** und nicht das experimentell vorgegebene, statisch-faktoriell manipulierte Informations-Angebot in den Mittelpunkt der Forschung stellen.

2.2.5.3. Zuverlässigere Diagnose von Unternehmensentwicklungen mit "Schematischen Gesichtern"?

Die Effizienzprüfung Schematischer Gesichter richtet sich auf die Frage, ob dieses graphische Hilfsmittel eine bessere Klassifikation zukünftiger Unternehmensentwicklungen ermöglicht als eine Darbietung numerischer Informationen. Man vergleicht die Klassifikationsgüten von Versuchspersonen, die Schematische Gesichter benutzen, mit solchen, die eine kleine Menge von Kennzahlen oder ausgewählten numerischen Daten aus Bilanz (i.e.S.) und Erfolgsrechnung erhalten. Das komplexe Beurteilungsproblem "Bilanzanalyse" wird somit auf ein Problem der adäquaten Gewichtung weniger vorgegebener Schlüsselinformationen reduziert. Es handelt sich um eine ähnliche Problemstellung wie bei den sog. "Linsen"-Studien, die im Zusammenhang mit der Darstellung unseres Effizienzkonzeptes gewürdigt werden.⁷¹⁾

Die hierzu durchgeführten Studien von Moriarity und Stock/Watson weisen auf eine **mäßige, positive Wirkung Schematischer Gesichter** hin (vgl. Abbildung 2.8.): Sowohl Studenten als auch Wirtschaftspraktiker erreichen signifikant höhere Klassifikationsgüten mit Schematischen Gesichtern als mit Kennzahlen. Die Effekte sind jedoch nicht besonders stark ausgeprägt: Nach unserer

Abbildung 2.8.: Einflüsse graphischer Hilfsmittel auf die Effizienz betrieblicher Entscheidungen (Teil 2)

- Schematische Gesichter vs. Tabellen -

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Moriarty (Graphics 1979)	Beurteilung, ob ein Unternehmen im folgenden Jahr insolvent wird	277 Stud.	Schematische Gesichter a) ohne, b) mit Erklärung, c) 10 ausgewählte Daten aus Bilanz und Erfolgsrechnung, d) 13 ausgewählte Kennzahlen; jeweils für sechs Jahre vor Insolvenz, 22 zu klassifizierende Firmen	Anzahl der korrekt klassifizierten Firmen	Sign. Untersch. zw. allen vier Bed. (Re-Analyse "Table 4", erkl. Var. 3,0 v.H.) aber nur Untersch. a), b) vs. d) sign., beruht nach Moriarty S. 214 f. auf geringem Sachverstand der Vpn, nicht auf Schemat. Gesichtern
		20 Manager	Schematische Gesichter mit Erläuterung vs. Kennzahlen, jeweils sechs Jahre vor Insolvenz, 11 zu klassifizierende Firmen	Anzahl der korrekt klassifizierten Firmen	mit Gesichtern sign. weniger falsch klassifiz. Firmen (Re-Analyse "Table 7", Chiquadratstest, tetrach. Kor.: 0,25)
Stock/Watson (Graphics 1984)	Beurteilung, ob ein Bond-Rating im nächsten Jahr steigt, fällt oder gleich bleibt	123 Stud. 14 Manager	Schematische Gesichter vs. sechs Kennzahlen, jeweils sechs Jahre vor neuem Rating, 21 zu klassifizierende Firmen	Anteil der korrekt klassifizierten Veränderungen bzw. Konstanz	sign. weniger Fehler bei Schematischen Gesichtern (Re-Analyse F-Wert S. 199: erkl. Var.: 7,9 v.H. keine sign. Interaktion Stud./Manager)

- Baumstrukturen vs. Tabellen -

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Hauschildt (Graphische Unterstützung 1985)	Bilanzanalysen von Krisenunternehmen	103 Stud. je vier Fälle	drei Bed.: a) Tabelle, b) Baumstruktur nur während Analyse, c) Baumstruktur während Analyse und Feedbackveranstaltung	Treffsicherheit Erfolgslage, gewichtete Maßgröße zur Analysequalität	mit Baum sign. höhere Treffsicherheit, mit Baum in Feedback-Veranst. sign. schnellerer Lernfortschritt

- mono- vs. multichromatische Darstellung -

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Benbasat/Dexter (Graphical Information Presentation 1985)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verschiedene Verkaufsgebiete	35 Stud.	ohne vs. mit Farbe (zweite exp. Bed. Tabelle vs. Graphik, 2x2-Design, außerdem Kognitiver Stil: Group Embedded Figures Test)	Erzielte Gewinne, relativ zum Optimum	Farbe sign. besser, sign. Interaktion: Feldabhängige mit Farbe stärker verbessert, keine sign. Interaktion mit Präsentationsform
Benbasat/Dexter/Todd (Graphical Information Presentation 1985)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verschiedene Verkaufsgebiete	66 Stud.	ohne vs. mit Farbe (zweite exp. Bed. Tabelle vs. Graphik, 2x2-Design)	Erzielte Gewinne, relativ zum Optimum	n. s., keine sign. Interaktion mit Präsentationsform
Benbasat/Dexter (Time Constraints 1985)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verschiedene Verkaufsgebiete	65 Stud.	ohne vs. mit Farbe (zweite exp. Bed.: Tab. vs. Graphik vs. beides, dritte exp. Bed. Zeitlimit 5 vs. 15 Min. 2x3x2-Design)	Erzielte Gewinne, relativ zum Optimum, ob 99 v. H. erreicht werden	Farbe sign. besser, (Re-Analyse von Table 3 a) und b) zusammengefaßt, Chiquadratstest) Kein Interaktionstest, aber unter 5-Minuten-Bedg. Farbe sign. besser, bei 5 Min. n.s.

Re-Analyse beträgt die mittlere erklärte Varianz ca. 5,6 v. H., wobei die erklärten Varianzen der Einzelstudien nur wenig um dieses Mittel schwanken.⁷²⁾

Anzumerken ist, daß in der Studie von Moriarity der Vergleich ausgewählter Daten aus Bilanz und Erfolgsrechnung mit Schematischen Gesichtern **keine** signifikanten Überlegenheit der Schematischen Gesichter zeigt. Überdies liefern auch die Bilanz- und Erfolgsrechnungsdaten signifikant bessere Klassifikationsgüten als die Kennzahlen (einseitiger Test). Moriarity begründet dieses mit mangelndem Sachverstand der studentischen Analytiker, die den Informationsgehalt der Kennzahlen falsch einschätzen. Man kann jedoch auch die Alternativhypothese aufstellen, daß die unverdichteten Daten einen höheren Informationsgehalt aufweisen als die Kennzahlen.

Bei Stock/Watson fehlt die Versuchsbedingung "Bilanz- und Erfolgsrechnungsdaten". Dafür gibt es bei ihnen eine Ergänzungsuntersuchung mit 11 weiteren Praktikern, denen zusätzlich die geschätzten Klassifikationen einer multiplen Diskriminanzanalyse für die Veränderung des Bond-Ratings zur Verfügung gestellt werden. Diese Zusatzinformation bewirkt – entgegen den Erwartungen – keine signifikante Verbesserung der Klassifikationsgüte.⁷³⁾

Bei beiden Studien ist kritisch zu fragen, ob es sich um einen "fairen" Vergleich gehandelt hat: Stock und Watson weisen explizit darauf hin, daß sie bei der Gruppe mit Schematischen Gesichtern den stärker wahrgenommenen "inneren" Gesichtszügen "Nasenlänge", "Augenwinkel" und "Mundkontur" die deutlich am stärksten diskriminierenden Kennzahlen zugewiesen haben.⁷⁴⁾ Dadurch wird der Gruppe mit Schematischen Gesichtern ein **impliziter Gewichtungsvorteil** verschafft. Man hätte der Gruppe mit Kennzahlen verbale oder numerische Hinweise auf die Ungleichgewichtigkeit der Kennzahlen geben müssen, um Chancengleichheit herzustellen und eine äquifunktionale Dekodierung der Information zu ermöglichen.

Auch bei den Schematischen Gesichtern wurde geprüft, ob das graphische Hilfsmittel den Zeitverbrauch reduziert. Während bei Moriarity studentische Analytiker mit Schematischen Gesichtern signifikant schneller zu einem Urteil kommen, können Stock/Watson keine signifikanten Unterschiede im Zeit-

verbrauch feststellen.⁷⁵⁾

Insgesamt gesehen steht ein fairer Test der Schematischen Gesichter gegenüber verbalen oder numerischen Gewichtungshilfen noch aus. Außerdem ist es für praktische Anwendungen nicht unproblematisch, eine komplexe Firmenbeurteilung auf ein Problem der angemessenen Gewichtung weniger vor-"gegebener" Schlüsselinformationen zu reduzieren. Schließlich spricht gegen Schematische Gesichter, daß der Adressat nur schwer erkennen kann, wie der Sender die Urteilsdimensionen gewichtet. Wenn der Adressat nur ein Schematisches Gesicht erhält, hat er es sehr schwer, eine alternative Gewichtung anzuwenden. Als exklusive Information könnten Schematische Gesichter erhebliche Akzeptanzprobleme aufwerfen: Zur Illustration stelle man sich vor, daß die Ergebnisse von Warentests nur als Schematische Gesichter publiziert würden. Dies schließt nicht aus, daß Schematische Gesichter eine nützliche Zusatzinformation darstellen könnten.

Offen ist die Frage, worauf die bisher ermittelte (mäßige) Überlegenheit Schematischer Gesichter beruht. Wir sehen zwei Ursachen:

a) angemessenere Einschätzung des Verhältnisses von Reiz zu prognostizierendem Ereignis, weil im Schematischen Gesicht die prognoserelevanten Reize graphisch stärker gewichtet werden.

b) konsistentere Gewichtung der einzelnen Reize bei verschiedenen Beurteilungsproblemen, weil die bildhafte Urteilsintegration gleichmäßiger erfolgt als die Verknüpfung numerischer Daten.

Mit Hilfe eines "lens-model-designs" könnte man die relative Bedeutung beider Erklärungen systematisch überprüfen. Man könnte auch feststellen, wie stark bestimmte Gesichtszüge von den Analytikern gewichtet werden.

2.2.5.4. Unterstützung der Bilanzanalyse durch Baumstrukturen

Zur Wirkung von Baumstrukturen auf die Effizienz komplexer betriebswirtschaftlicher Beurteilungsprozesse ist uns nur die Studie von Hauschildt bekannt. Sie zeigt eine signifikant höhere Treffsicherheit der Erfolgslage bei baumstrukturartig angeordneten numerischen Informationen; außerdem einen signifikant schnelleren Lernfortschritt, falls die Baumstruktur auch während der Feedbackveranstaltung benutzt wird. Im Rahmen unserer eigenen Arbeit werden wir am gleichen Material wie Hauschildt erforschen, auf welchen Pfaden diese Effizienzwirkungen zustandekommen. Die Bewertung des Baumstruktureffektes wird daher bis zum vierten Teil dieser Schrift zurückgestellt.

Anzumerken ist an dieser Stelle, daß einige Experimente zur systematischen Gestaltung von Bildschirmmasken eine signifikante Reduktion von Zeitverbrauch und psychischer Belastung durch problemgerecht strukturierte Zeichenanordnungen belegen.⁷⁶⁾

2.2.5.5. Erste Anzeichen für positive Wirkungen mehrfarbiger Informationsangebote

Die Forschungsergebnisse zur Wirkung farbiger Darstellungen sollen nach den bei Ives, DeSanctis, Benbasat und Kollegen herangezogenen Sammelreferaten uneinheitlich sein und stark von den Anforderungen der Aufgaben abhängen. Die drei uns bekannten Studien von Benbasat und Kollegen, in denen die Wirkungen farbiger Darstellungen auf uns interessierende Effizienzmaße geprüft werden, zeigen demgegenüber eine positive Tendenz: In zwei der drei Studien zeigt sich ein signifikant positiver Einfluß (vgl. Abbildung 2.8.). Hinsichtlich des Zeitverbrauches ergibt sich in einem der drei Experimente ein signifikant negativer Einfluß. Aus diesen Ergebnissen sollte man noch keine Verallgemeinerungen vornehmen. Man sollte vielmehr abwarten, ob sich die positiven Effekte in zukünftigen Studien bestätigen lassen. Hierfür spricht auch die differenzierte Erwartung von Benbasat und Ko-Autoren, die nur unter bestimmten Bedingungen mit positiven Wirkungen farbiger Informationsangebote rechnen.⁷⁷⁾

2.2.6. Effizienzsteigerung bei relevantem Feedback

Daß man mit Feedback-Informationen Lernprozesse stimulieren und die Qualität zukünftiger Entscheidungen verbessern kann, ist eine Grundannahme des Controlling. Die Untersuchungen von Mock und Ko-Autoren **bestätigen** diese Hypothese (vgl. Abbildung 2.9.).

In einem computergestützten Unternehmensspiel zur Abstimmung von Produktions- und Absatzmengen weist Mock nach, daß unverzügliches Feedback zu signifikant höheren Spielgewinnen und signifikant niedrigeren Kosten führt. In Nachfolgestudien belegen Mock und Ko-Autoren, daß vollständiges und detailliertes Feedback über Budgetabweichungen zu signifikant besseren Spielergebnissen führt. Die erklärten Varianzen für die Effizienzgrößen liegen zwischen 3,7 und 17,9 v. H., die mittlere Varianzerklärung beträgt ca. 8,6 v. H.⁷⁸⁾ Zusätzliche Auswertungen zeigen unter Feedback-Bedingungen größere Lernfortschritte von Entscheidungsqualität und Zeitverbrauch.⁷⁹⁾

Bei den Untersuchungen von Mock und Mitarbeitern kann man aufgrund der Konstruktion des Spieles, für das eine optimale Lösung existiert, davon ausgehen, daß das Feedback die Versuchspersonen zu einer besseren Lösung befähigt. Dies ist bei den nachfolgend referierten Studien nicht immer der Fall und könnte die teilweise abweichenden Befunde erklären:⁸⁰⁾

* In der Arbeit von Putz-Osterloh erhält eine experimentelle Gruppe eine positive Rückmeldung, daß ihre Leistung im "oberen Drittel" läge, **unabhängig** von der tatsächlichen Leistung. Die Autorin will untersuchen, wie Mißerfolgserlebnisse verarbeitet werden. Sie geht davon aus, daß positive Rückmeldungen das Selbstbewußtsein stärken und die Leistungsbereitschaft und -fähigkeit erhöhen. Jedoch zeigt sich keine Beziehung zum Effizienzmaß "Anzahl der Perioden mit steigendem Zahlungsmittelbestand", das in ihrem Unternehmensspiel verwendet wird.

* In der Studie von Kessler/Ashton sind wenige, angebotene Kennzahlen zu einem Bond-Rating zu verknüpfen, das dem von Moody's möglichst nahekommt. Es handelt sich um ein in der psychologischen Forschung häufiger

Abbildung 2.9.: Einflüsse von Feedbackinformationen auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Mock (Information Structures 1969)	Unternehmensspiel Produktion und Absatz	21 Stud. 25 Manager 26 Doktorand.	verzögerter vs. unmittelbarer Feedback	1. Gewinne 2. Kosten	1. sign. höhere Gewinne bei unmittelb. Feedback, einseitig. Test Korr.: 0,21 2. sign. geringere Kosten bei unmittelb. Feedback, Korr.: 0,34 (Re-Analyse von Tab. 6 und 9, punkt-bis-seriale Korr.)
Mock (Budget Information 1973)	Unternehmensspiel, Produktion und Absatz	40 Stud.	ohne vs. mit Feedback über Budgetabweichung (Budget war von Vpn festzulegen)	1. Gewinne 2. Kosten	1. sign. höhere Gewinne mit Feedback, eins. Test Korr.: 0,26 2. sign. geringere Kosten mit Feedback, Korr.: 0,34 (Re-Analyse von Tab. 1 punkt-bis. Korr.)
Driver/Mock (Experimental Results 1976)	Unternehmensspiel, Produktion und Absatz	54 Stud.	nur Soll- und Ist-Größen vs. zusätzlich absolute u. rel. Budgetabweichung. Abweichung kann bei erster Bed. auch errechnet werden	Gewinne ("average profits")	höhere Gewinne mit vollständigem Feedback, Signifikanztests nur für Subgruppen, dort sign.
Kessler/Ash-ton (Feedback 1981)	Bond-Rating	69 Stud.	4 Feedback-Bed.: a) nur Trefferquote, b) zusätzl. Korr. Kennzahlenwert mit eigenem Rating, c) nur Korr. Kennzahlenwert mit Moody's Rating, d) alle Arten von Feedback	Übereinstimmung des abgegebenen Ratings mit dem von Moody's: 1. Korrelationskoeffizient 2. absolute Abweichung	1. sign. höhere Übereinstimmung, sign. höherer Lernfortschritt bei c) und d) 2. kein sign. Einfluß des Feedback-Typs
Putz-Osterloh (Problemlöseleistung 1983)	Unternehmensspiel, Produktion und Absatz	90 Stud.	Gruppe mit positiver Rückmeldung, daß ihr Erfolg im oberen Drittel läge (unabhängig vom tatsächlichen Erfolg) vs. übrige Gruppen	Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	n.s. (Re-Analyse F-Wert, S. 108: erkl. Varianz 1,0 v. H.)
Jacoby/Mazursky/Troutman/Kuß (Feedback 1984)	Aktienkauf, simuliert mit historischen Kursen	17 Finanzanalysten	Beschaffung von Feedback-Informationen über Kursentwicklung im vergangenen Quartal, zw. Kursentwicklung im vergangenen und lfd. Quartal besteht eine starke negative Beziehung	Kursentwicklung pro Periode, bereinigt um Dividenden	negativ sign.
Mock/Vasarihelyi (Cognitive Style 1984) Experiment 154	Unternehmensspiel, Produktion und Absatz	130 Stud.	a) Vollständigkeit b) Feinheit ("coarseness") des Feedback über Budgetabweichungen	Gewinne ("average profits")	a) sign. höhere Gewinne bei vollst. Feedback, erkl. Varianz: 3,7 v.H. b) sign. geringere Kosten bei feinem Feedback, erkl. Varianz: 5,7 v.H. Tend. sign. Interaktion (Re-Analyse von "Table 5", S. 73)

2.2 untersuchtes Problem des Erlernens probabilistischer Beziehungen, diesmal eingebettet in einen betriebswirtschaftlichen Kontext. Wenn man den Versuchspersonen nur mitteilt, wie oft sie das Rating getroffen haben ("outcome feedback"), ist der Lernfortschritt signifikant schlechter als wenn man ihnen mitteilt, wie die Informationen aufgrund eines statistischen Modells bei Moody's Rating durchschnittlich gewichtet werden ("cognitive feedback"). Das Experiment bestätigt die psychologische Theorie, daß reines Feedback über Ergebnisse einer Instruktion über **Zusammenhänge** unterlegen ist.

* Ein **negativer** Zusammenhang zeigt sich in der Studie von Jacoby/Mazursky/Troutman/Kuß: Finanzanalysten, die sich besonders häufig über die Entwicklung der Kurse des vergangenen Quartals informieren, erzielen signifikant geringere Renditen. Dies kann damit erklärt werden, daß ein Entscheidungsträger, der die Kursentwicklung in der Folgeperiode nur aus dem Kurs der Vorperiode ableitet, in der nachgespielten historischen Situation einen Verlust von 41,7 % realisiert.

Wir können folgendes Ergebnis festhalten: **Feedback wirkt dann effizienzsteigernd, wenn die Feedback-Information eine positive Beziehung zum Effizienzkriterium aufweist.** Fraglich ist, ob man diese positive Beziehung stets so leicht bestimmen kann, wie in den referierten experimentellen Studien von Kessler/Ashton sowie Mock und Ko-Autoren. Die Arbeiten weisen ferner darauf hin, die Wirkung von Feedback differenzierter zu erfassen und dem Feedback über **Zusammenhänge** eine stärkere Beachtung zu schenken. Erforscht werden sollte auch, unter welchen Bedingungen die in manchen psychologischen Studien ermittelten **negativen** Wirkungen eines reinen **Ergebnis-Feedbacks** bei komplexen betriebswirtschaftlichen Problemen zu erwarten sind.⁸¹⁾ Kritisch ist auch zur Feedback-Forschung anzumerken, daß nicht untersucht wurde, **wie** sich die Veränderung des Informationsverhaltens vollzieht, auf der die festgestellten Lernfortschritte beruhen.

2.2.7. Formale Entscheidungshilfen: noch wirksamer bei angemessenem Informationsverhalten

Die Entwicklung formaler Entscheidungsmodelle zur Erklärung und Gestaltung betriebswirtschaftlicher Probleme weist eine beachtliche Tradition auf, die durch die Verbreitung leistungsstarker Rechner einen starken Aufschwung und Anwendungsbezug gefunden hat. Die Leistungen solcher Entscheidungshilfen zur Verbesserung betriebswirtschaftlicher Entscheidungsprobleme sind unstrittig. Eine systematische Bestandsaufnahme dieser Leistungen wird nicht angestrebt. Wir wollen lediglich an ausgewählten, nicht-repräsentativen Beispielen illustrieren, daß sich noch größere Verbesserungen der Entscheidungsqualität ergeben können, wenn man das Informationsverhalten der Benutzer bei der **Anwendung** der Modelle berücksichtigt. Ausgeschlossen sind somit auch verhaltenstheoretische Überlegungen, die mit der Modellformulierung und -implementierung zusammenhängen.⁸²⁾

Eine Berücksichtigung des menschlichen Informationsverhaltens scheint uns beispielsweise dann zweckmäßig, wenn:

- * geprüft werden soll, ob die Prämissen, die einer Entscheidungshilfe zugrundeliegen, erfüllt sind oder beurteilt werden soll, in welchem Maße ihre Verletzung toleriert werden kann,
- * eine Auswahl zwischen mehreren Algorithmen oder Heuristiken vorzunehmen ist oder mehrere Verfahren in unterschiedlicher Weise kombiniert werden können,
- * eine Startlösung von einem Benutzer spezifiziert werden soll, die Ablauf und Ergebnis des Verfahrens beeinflusst,
- * der Benutzer über einen Abbruch des Verfahrens entscheiden soll, sei es wegen mangelnder oder zunehmend langsamerer Konvergenz auf eine Optimal-lösung hin,

- * der Benutzer eine Gewichtung mehrerer Ziele innerhalb pareto-optimaler Lösungen vornehmen soll, oder wenn der Benutzer bei Entscheidungen unter Unsicherheit Risikopräferenzen spezifizieren soll.

Die in Abbildung 2.10. zusammengefaßten experimentellen Effizienzprüfungen formaler Entscheidungshilfen geben einige Hinweise auf die empirische Gültigkeit dieser Thesen. In allen fünf Experimenten ergibt sich in der Versuchsbedingung "mit" Entscheidungshilfe(n) eine signifikant höhere Effizienz. Uns interessiert die Frage, welche Rolle das Informationsverhalten bei diesem Befund spielt. Hierzu kann folgendes berichtet werden:

1. In der Arbeit von Benbasat geht es um die Optimierung der Beschaffung und Beständehaltung.⁸³⁾ Als Entscheidungshilfen werden Bedarfsprognosen, Bestellmengenauflösung und ein Bestellmengenvorschlag, basierend auf einem Algorithmus von Hadley und Whitin, angeboten. Den Entscheidungsträgern wird mitgeteilt, daß der Algorithmus auf vereinfachenden Annahmen beruhe, die im Experiment nicht erfüllt seien, so z. B. die Normalverteilung der Nachfrage und daß er keine Retouren ("back-orders") berücksichtige. Ein Kontrolllauf, der nur den Vorschlägen der Entscheidungshilfen folgt, ergibt bessere Ergebnisse als bei den meisten der 32 Entscheidungsträger. Nur drei Versuchspersonen, die Entscheidungshilfen und Graphiken erhalten, sind in der Lage, die Verletzung der Prämissen zu erkennen und problemgerecht von den Empfehlungen des Algorithmus abzuweichen.

2. In einer weiteren Studie zum gleichen Entscheidungsproblem erhält eine Experimentalgruppe die Möglichkeit, ein **Simulationsprogramm**⁸⁴⁾ einzusetzen, das die hypothetischen Auswirkungen der zugrundegelegten Entscheidung ermittelt. Es zeigt sich, daß sowohl feldunabhängige als auch feldabhängige Entscheidungsträger ihre Leistung mit diesem einfachen Instrument deutlich verbessern.⁸⁵⁾ Interessant ist, daß sich die Leistungsunterschiede zwischen beiden Gruppen verringern, so daß die vorher signifikant schlechteren feldabhängigen Personen mit den feldunabhängigen gleichziehen. Die Autoren führen dies darauf zurück, daß die Entscheidungshilfe spezifische Defekte im Informationsverhalten der feldabhängigen Personen überwindet.⁸⁶⁾

Abbildung 2.10.: Einflüsse formaler Entscheidungshilfen auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Informationsbedingung	Effizienzgröße	Befund
Benbasat (Effects 1974)	Unternehmens- spiel, Beschaf- fung und Pro- duktion	32 Stud.	ohne vs. mit Entscheidungs- hilfen (Bedarfsprognose, Bedarfsauflösung und Be- stellmengenanschlag), and. exp. Bed.: Angebots- menge, Tabelle vs. Graphik, "exception reporting"; ferner: Kognitiver Stil und Vorwissen	Kosten	sign. geringere Kosten mit Hilfen, tend. sign. Inter- aktion von Graphik und Entscheidungs- hilfen, andere In- teraktionen n.s.
Uebele (Planungs- techniken 1980)	Unternehmens- spiel, Preis- entscheidun- gen	80 Stud.	mit vs. ohne (formalisierte) Entscheidungsbaume	Gewinnzuwächse (S. 236)	sign. höhere Ge- winnzuwächse mit Entscheidungs- baum
Benbasat/ Dexter (In- dividual Dif- ferences 1982)	Unternehmens- spiel, Beschaf- fung und Pro- duktion	61 Stud.	ohne vs. mit Simulations- modell	Deckungsbei- träge	Sign. Test getrennt nach Stilen, beide Male mit Hilf sign. besser (einseitiger Test)
McIntyre (Marketing Models 1982)	simul. Verkaufs- einsatzplanung, Aufteilung des Budgets auf ver- kaufsgebiete	69 Stud.	ohne vs. mit Entscheidungs- modell CALLPLAN; and. exp. Bed.: niedrige vs. hohe Zufallsschwankungen, drei vs. sechs Verkaufsgebiete; ferner: Kognitive Stile	Erzielte Gewin- ne, relativ zum Optimum	mit Modell sign. höhere Gewinne, keine sign. Inter- aktionseffekte
Leichtfuß (Kapital- budgetie- rung 1984)	Unternehmens- spiel, Abstimmg. von Investition und Finanzierung	64 Stud.	Computerkonferenz vs. Compu- terkonferenz mit Unterstüt- zung durch einen Dekomposi- tionsalgorithmus vs. rein algorithmische Lösung	Endwert des di- visionalisierten Unterneh- mens	Rein algorithm. Lösung und modell- gest. Lösung sign. besser als reine Konferenz

3. McIntyre untersucht die Effizienzwirkung des CALLPLAN-Modells zur Unterstützung der Verkaufseinsatzplanung. Versuchspersonen mit Entscheidungsmodell erzielen signifikant höhere Gewinne, machen weniger grobe Fehler und zeigen schnellere Lernfortschritte. Die Detailanalyse zeigt, daß man in 57 von 84 Fällen, in denen von der Modellempfehlung abgewichen wird, ein höherer Gewinn bei modellkonformem Handeln möglich gewesen wäre. Die meisten Abweichungen beruhen darauf, daß das Modell eine stärkere Veränderung des status quo empfiehlt. Dies kann man jedoch nur dann als "Fehlverhalten" ansehen, wenn die experimentelle Modellprämisse einer stationären stochastischen Abhängigkeit gegeben ist. Bei realen Problemen könnte die Diskrepanz eine effizienzsteigernde Ursachenforschung anregen.

4. Uebele erforscht die Wirkungen von Entscheidungsbäumen bei Preisentscheidungen. In jeder der acht Spielperioden haben die Versuchspersonen die Wahl zwischen Beibehaltung, Erhöhung oder Senkung des bisherigen Preises. Nach jeder Entscheidung werden sie über die Umweltreaktionen, die abgesetzte Menge und den erzielten Erfolg informiert. In der Experimentalgruppe haben sich die Versuchspersonen vor der 5. Entscheidung mit dem Entscheidungsbaum vertraut zu machen und einen **selbst** entworfenen Baum als Entscheidungsgrundlage zu benutzen. Ab der sechsten Periode ist die Verwendung eines **standardisierten** Entscheidungsbaumes, versehen mit Erläuterungen eines fiktiven Stabes, obligatorisch. Uebeles Effizienztest zeigt, daß die kumulierten Gewinne der beiden Gruppen sich nach der vierten Periode nicht signifikant unterscheiden. Nach Periode acht hat die Gruppe mit Entscheidungsbaum jedoch einen signifikant höheren Gewinn erreicht. Bemerkenswert ist, daß sich die höheren Leistungen der Experimentalgruppe in einer **signifikanten Verbesserung des Informationsverhaltens** niederschlagen: Die Mitglieder der Experimentalgruppe schätzen den Wert einer zusätzlich angebotenen Information und die Wahrscheinlichkeit, daß der Spielgewinn unter einen "Vorgabe-Wert" fällt, signifikant besser ein. Außerdem sind die Mitglieder der Experimentalgruppe mit ihrer Leistung signifikant zufriedener.

5. In der Arbeit von Leichtfuß wird die Kapitalbudgetierung in einem divisionalisierten Industrieunternehmen untersucht. Verglichen werden die Effizienzwirkungen dreier Konstellationen: a) reine Computerkonferenz von Mitgliedern der Zentrale und der Divisionen, b) Computerkonferenz, bei der ein

Dekompositionsalgorithmus Empfehlungen ausspricht, c) rein algorithmische Lösung. Bei der zweiten Bedingung können die Teilnehmer eine Startlösung festlegen und bestimmen, wieviele Iterationen durchlaufen werden sollen. Im Ergebnis zeigt die rein algorithmische Lösung zwar die höchsten Endwerte des Firmenwertes, aber auch die schlechtesten motivationalen Bewertungen. Die Computerkonferenz mit Entscheidungsunterstützung führt zu Endwerten des geplanten Projektportfeuilles, die nahe am Optimum liegen, beläßt aber den Entscheidungsträgern einen großen Entscheidungsspielraum und führt daher auch zu einem signifikant höheren Grad an wahrgenommener Entscheidungsautonomie als die rein algorithmische Lösung. Lösung. Beachtlich für unsere Thesen sind die **Ursachen** für das gute Abschneiden des heuristisch-algorithmischen Prinzips: Die hohen Vermögensendwerte beruhen nämlich bei den meisten Spielgruppen auf ausgezeichneten Startwerten. Sie ermöglichen es, den Entscheidungsprozeß mit nahe am Optimum liegenden Werten früher als in der algorithmischen Lösung abubrechen und in der Praxis u. U. recht kostspielige weitere Iterationen zu vermeiden. Dies belegt, daß die Einbeziehung des menschlichen Informationsverhaltens nicht nur aus motivationalen, sondern auch aus ökonomischen Überlegungen gerechtfertigt ist.

Auch wenn die dargestellten Experimente für die Forschung über Operations Research Modelle nicht repräsentativ sind, unterstreichen sie doch die Notwendigkeit der Einbeziehung verhaltenswissenschaftlicher Erkenntnisse. Wir teilen daher die Überzeugung von Schmidt: "... daß eine Integration verhaltenswissenschaftlicher Erkenntnisse in die betriebswirtschaftliche Entscheidungsforschung unter dem Effizienzgesichtspunkt unumgänglich ist."⁸⁷⁾ Eine solche Integration darf nicht nur von Forschern gefordert werden, die dem Operations Research nahestehen. Auch verhaltensorientierte Wissenschaftler sollten die Integration voranzutreiben, in dem sie sich stärker mit den Anforderungen beschäftigen, die formale Entscheidungshilfen an menschliches Informationsverhalten stellen.

2.3. PERSONLICHKEIT UND EFFIZIENZ BETRIEBLICHER ENTSCHEIDUNGSPROZESSE

2.3.1. Problemstellung

Persönlichkeitsmerkmale werden nicht nur in der psychologischen, sondern auch in der (amerikanischen) betriebswirtschaftlichen Forschung häufig zur Erklärung des Informationsverhaltens und dessen Effizienzwirkungen herangezogen. Vor allem für die Gestaltung von Management-Informationen- und Entscheidungs-Unterstützungs-Systemen wurde bereits Anfang der siebziger Jahre eine stärkere Beachtung von Persönlichkeitsvariablen, namentlich von unterschiedlichen "Kognitiven Stilen", gefordert:

"Managers need 'information' that is geared to THEIR psychology NOT to that of their designers."¹⁾

"The central assumption of our study is that there needs to be a fit between the decision maker's cognitive style and the information processing constraints of his task."²⁾

"The assumption here is that people differ in cognition but that different categories of people can be identified with similar thought processes. The approach here would be similar to the 'contingency theory' school in organization theory. (...) The advantage of this approach is that one can generalize yet be concerned with individual differences."³⁾

"Finally, the characteristics of the decision maker must become prime considerations in all areas of information systems development."⁴⁾

Nach einer recht intensiven "Jäger- und Sammlerphase", in der man Einflüsse verschiedener Stilmaße auf Informationsverhalten und Effizienz untersucht hat, ist die anfängliche Begeisterung einer gewissen Ernüchterung gewichen. Huber meint sogar, daß nicht nur die bisherige, sondern auch die zukünftige Forschung über Kognitive Stile wenig geeignet sei, operationale Handlungsanweisungen für das Design von Management-Informationen- und Entscheidungs-Unterstützungs-Systemen abzuleiten.⁵⁾ Seine Aussage stützt sich allerdings nicht auf eine eigene, systematische Bestandsaufnahme, sondern auf ausge-

wählte Sammelreferate, die er zum Beleg seiner These heranzieht.⁶⁾ wurden.

Wir führen eine solche systematische Bestandsaufnahme durch, um die Frage zu beantworten, ob die empirischen Ergebnisse den hohen **Effizienzanspruch** rechtfertigen.⁷⁾ Dies ist dann gegeben, wenn Entscheidungsträger mit bestimmten Persönlichkeitsmerkmalen entweder⁸⁾

1. **generell** effizienter sind als andere, oder wenn sie
2. bei bestimmten Arten von **Aufgaben** durchgängig effizienter sind als andere, oder wenn sie
3. bei bestimmten Konstellationen des **Informationssystems** durchgängig effizienter sind als andere.

Im ersten Fall könnte man bereits bei der Personalauswahl und Karriereplanung auf Personen mit bestimmten kognitiven Eigenschaften achten, im zweiten Fall wäre die Zuordnung von Personen zu Aufgaben betroffen, im dritten Fall die Gestaltung von Informationssystemen. Bei starken und reproduzierbaren Unterschieden könnte man sich in allen drei Fällen die mühselige Arbeit ersparen, das Informationsverhalten unterschiedlicher Typen prozeßbegleitend zu erheben und zu überprüfen, wie und **warum** die festgestellten Unterschiede zustande kommen.

Wir bestimmen daher, wie häufig kognitive Persönlichkeitsmerkmale signifikante Haupteffekte und/oder signifikante Interaktionseffekte mit Merkmalen der Aufgabe oder des Informationssystems aufweisen.⁹⁾ Dabei bilden wir zwei Schwerpunkte:

- * Einen ersten Schwerpunkt bildet das schillernde, sehr unterschiedlich interpretierte und operationalisierte Konstrukt **"Kognitiver Stil"**, das in der amerikanischen Managementforschung besonders stark beachtet wird.
- * Einen zweiten Schwerpunkt bildet das in gängigen Tests erfaßte Konstrukt **"Intelligenz"**. Seine Wirkungen auf die Qualität komplexer Managementaufgaben sind in der jüngeren deutschsprachigen psychologischen Problemlö-

seforschung im Rahmen computergestützter Simulationen untersucht worden. Ergänzend werden Studien berücksichtigt, in denen man den Einfluß quantitativer Fähigkeiten analysiert.

Beiden Konzepten liegt die Idee zugrunde, daß man mit vergleichsweise einfachen, standardisierten Papier-und-Bleistift-Tests subjektabhängige Prädispositionen messen kann, Probleme in einer bestimmten Art und Weise wahrzunehmen und zu bearbeiten. Während die Forschungsrichtung "Kognitive Stile" meist eine Kontingenz-Hypothese bezüglich der Merkmale eines Informationssystems vertritt, ist der generelle Effizienzanspruch der Intelligenzforschung ungebrochen. Erst in neueren Studien der deutschsprachigen Problemlöseforschung werden Kontingenz-Hypothesen formuliert und überprüft.

Selbstverständlich kann unsere Schwerpunktbildung nicht beanspruchen, die Erklärungskraft des weitaus reichhaltigeren Konstruktes "Persönlichkeit" auszuschöpfen. Man kann jedoch davon ausgehen, daß wir die kognitiven Eigenschaften erfaßt haben, die besonders häufig experimentellen Effizienzprüfungen unterzogen werden.

2.3.2. Die Grundtendenz: Weder "Kognitive Stile" noch "Test-Intelligenz" führen generell zu besseren Leistungen

Das Ergebnis unserer Bestandsaufnahme ist enttäuschend (zur Dokumentation der Bewertung der Studien vgl. die Abbildungen in den nachfolgenden Kapiteln):

1. **Kognitive Stile:** Von 41 geprüften Haupteffekten sind nur 12 signifikant, davon widersprechen sich zwei durch richtungsconträre Befunde, d. h.: 73 v. H. falsifizieren die Haupteffekt-Hypothese. Von 21 getesteten Interaktionseffekten sind nur 2 signifikant, d. h.: 90 v. H. falsifizieren die Interaktionseffekt-Hypothese.

2. **Test-Intelligenz und quantitative Fähigkeiten:** Von 24 geprüften Haupteffekten sind nur 2 signifikant, d. h.: **92 v. H. falsifizieren die Haupteffekt-Hypothese.** Von 7 getesteten Interaktionseffekten sind nur 2 signifikant, d. h.: **71 v. H. falsifizieren die Interaktionseffekt-Hypothese.**

Dieses Ergebnis steht in einem krassen Widerspruch zum hohen Effizienzan-spruch beider Forschungsrichtungen. Es fällt noch schlechter aus als kritische Bewertungen in Sammelreferaten anderer Forscher¹⁰⁾ und mahnt Apologeten vor einer allzu unkritischen Rezeption.¹¹⁾

Bemerkenswert ist auch die **Art** des Ergebnisses: Im Gegensatz zur Kontingenz-Hypothese sind bei den Kognitiven Stilen besonders wenige Interaktionseffekte signifikant. Im Gegensatz zur generellen Effizienzvermutung fallen bei den Tests der "Test"-Intelligenz besonders wenige Haupteffekte signifikant aus.

Bevor wir auf die Ursachen dieses Ergebnisses näher eingehen, sollen einige **Einwände** gegen unsere Art der Ergebnisbestimmung gewürdigt werden.

- a) Wenn man nur die Signifikanz von Ergebnissen aus zählt, ist auf die **Teststärke** zu achten. Bei Studien mit kleinem Stichprobenumfang und mäßigen Effekten in der Grundgesamtheit kann die Wahrscheinlichkeit für einen sog. "Fehler zweiter Art", d. h. die irrtümliche Annahme, daß **kein** Zusammenhang besteht, sehr hoch ausfallen. Sie nimmt unter diesen Bedingungen mit wachsender Zahl ausgewerteter Studien nicht etwa ab, sondern konvergiert gegen Eins.¹²⁾
- b) Die beobachteten Effekte fallen wegen **Meßfehler** der Persönlichkeitsvariablen geringer als die "wahren" Effekte aus.¹³⁾
- c) Die untersuchten Populationen von Studenten und Managern sind nicht repräsentativ für die Gesamtbevölkerung. Wegen einer **Stützung des Wertebereiches** der kognitiven Merkmale muß man mit schwächeren Zusammenhängen als in der Gesamtbevölkerung rechnen.¹⁴⁾

Zu a) Mangelnde Teststärke:

Gliedert man die Befunde nach dem **Stichprobenumfang**, zeigt sich folgendes Ergebnis (nur Haupteffekte, Signifikanzniveau 5 %, zweiseitiger Test):

Abbildung 2.11.: Differenzierung der festgestellten Einflüsse von kognitiven Persönlichkeitsmerkmalen nach dem Stichprobenumfang der Untersuchungen

	signifikante Unterschiede		nicht signif. Unterschiede	
Stichprobenumfang	Anzahl	v. H.	Anzahl	v. H.
unter 50 Fälle	7	23	23	77
genau 50 Fälle	2	20	8	80
über 50 bis 99 Fälle	2	10	19	90
100 und mehr Fälle	1	25	3	75

Die Auswertung kann unsere Vorbehalte nicht entkräften. Der Anteil **nicht** signifikanter Studien ist bei **allen** Stichprobenumfängen **sehr hoch**.

Zu b) Reliabilität der Messung:

Die **Reliabilität** der verwendeten Meßinstrumente erfüllt im allgemeinen die gängigen Anforderungen der klassischen Testtheorie. So ermittelt Fink durchaus akzeptable Werte für Cronbachs Alpha, die denen anderer Untersuchungen in etwa entsprechen (vgl. Abbildung 2.12).¹⁵⁾ Nach Fink sollen lediglich bei Drivers "decision style" die Test-Retest-Reliabilitäten zwischen 0,26 und 0,64 und damit unter generell akzeptierten Schwellenwerten liegen.¹⁶⁾ Auch wenn unsere Auswertung für diese Stilvariable schlechte Ergebnisse zeigt, kann dieses Einzelergebnis nicht unseren enttäuschenden Gesamtbefund erklären, denn nur 3 der 51 ausgewerteten Haupteffekte werden mit diesem Instrument durchgeführt. Die gelegentlich recht pauschale Kritik an der mangelnden Reliabilität der für kognitive Persönlichkeitsmerkmale entwickelten Meßinstrumente ist somit weder empirisch gerechtfertigt, noch

geeignet, das Ausbleiben signifikanter Zusammenhänge mit Effizienzgrößen zu erklären.¹⁷⁾

Abbildung 2.12.: Reliabilität häufig verwendeter Instrumente zur Messung Kognitiver Stile (nach Fink, Kognitive Stile 1985, S. 155)

	Pretest- Population (N=52)	Versuchs- Population (N=57)
a) Feldunabhängigkeit	0,81	0,89
b) Myers-Briggs-Type-Maße		
Extra-Intro-Version	0,75	0,79
Sensing-Intuition	0,74	0,73
Thinking-Feeling	0,86	0,93
Judging-Perceiving	0,81	0,81
c) Differenzierung REP-Test	0,66	0,73
d) Diskriminierung REP-Test	0,82	0,83
e) Analytisch-Heuristisch, Fragebogen von Barkin	0,69	0,80

Zu c) Gestutzte Stichproben

Für das Argument, daß die festgestellten Zusammenhänge niedriger als in der Gesamtbevölkerung ausfallen, gibt es gewisse Belege: Die Kognitiven Stile der Untersuchungspopulationen decken wegen berufsgruppenspezifischer Abhängigkeiten nur einen Teilbereich des gesamten Stilspektrums ab.¹⁸⁾ Studentische Populationen weisen bei Intelligenztests im allgemeinen höhere

Werte auf als der Bevölkerungsdurchschnitt.¹⁹⁾ Wenn man die Meßwerte solcher Stichproben mit Effizienzmeßwerten korreliert oder eine Medianteilung anhand der **Stichprobenmeßwerte** vornimmt, dann können sich geringere Zusammenhänge als in der Gesamtbevölkerung ergeben. Gewisse Hinweise auf die Gültigkeit dieser Vermutung geben Funke und Mock/Estrin/Vasarhelyi: Funke schließt aufgrund einer Voruntersuchung durchschnittlich test-intelligente Personen aus.²⁰⁾ Mock/Estrin/Vasarhelyi beschränken sich auf Personen, die in einer nachträglichen Klassifikation als "eindeutig" analytisch oder heuristisch klassifiziert werden können.²¹⁾ Das Ergebnis sind signifikante bzw. tendenziell signifikante Unterschiede in der ökonomischen Effizienzvariablen. Aber: In der Realität ist die Gruppe mit durchschnittlichen Meßwerten am umfangreichsten. Bei **betriebswirtschaftlichen Anwendungen** müssen wir damit rechnen, daß Middle- und Top-Manager überdurchschnittliche kognitive Fähigkeiten aufweisen. Die Stützung der Stichproben ist demnach kein methodisches Artefakt, sondern Reflex einer entsprechenden Verteilung der Merkmale in der Grundgesamtheit der hier relevanten Zielgruppe.²²⁾

Ingesamt gesehen, können formale Einwände das Ergebnis unserer Bestandsaufnahme nicht entkräften. Es ist sogar zu befürchten, daß die praktische Relevanz noch geringer ist als es unser Befund ausweist: Erstens besteht eine Tendenz, "signifikante" Ergebnisse eher zu publizieren als nicht signifikante, zweitens ist die statistische Signifikanz ein Akzeptanzkriterium, kein Relevanzkriterium für die Stärke einer Beziehung. Angaben über Effektstärken oder Bewertungen der praktischen Relevanz stellen jedoch eher die Ausnahme als die Regel dar.

2.3.3. Einflüsse Kognitiver Stile

Die zu Effizienzprognosen herangezogenen Kognitiven Stile können in vier Gruppen eingeteilt werden:²³⁾

1. **Perzeptionsbezogene Stilmaße** wie der "Group Embedded Figures Test" von Witkin zur Messung der **Feld(un)abhängigkeit**²⁴⁾ und der "Familiar Figures Test" von Kagan zur Bestimmung von **Impulsivität und Reflexivität**.²⁵⁾

2. Die auf der **Persönlichkeitstheorie von Jung** aufbauenden **"Myers-Briggs Type Indicators"**, die in der amerikanischen Managementforschung besonders stark beachtet werden.²⁶⁾

3. Verschiedene, von amerikanischen Betriebswirten entwickelte Meßverfahren, mit denen das Konstrukt **"analytisch/heuristisch"** operationalisiert werden soll.²⁷⁾

4. Eine Restgruppe, zu der die **"Kognitive Komplexität"**²⁸⁾, die zweidimensionalen Typenbildungen von Rowe und Driver²⁹⁾ und die **"Hemisphärendominanz"**³⁰⁾ zugeordnet werden.

Diese vier Gruppen von Stilmerkmalen weisen durchaus unterschiedliche Beziehungen zur Effizienz aus. Mit ihnen wollen wir uns jetzt befassen, um die Ursachen für den ungünstig ausgeprägten Gesamtbefund zu erforschen.

2.3.3.1. Perzeptionsbezogene Stile: Effizienzerklärung bei stilgerechten Aufgaben

Die perzeptionsbezogenen Stilmaße schneiden deutlich besser als die übrigen Kognitiven Stile ab (vgl. Abbildung 2.13.):

- * fünf Studien belegen signifikant positive Unterschiede,
- * vier Arbeiten zeigen tendenziell signifikant positive Einflüsse,
- * bei drei Untersuchungen ergibt sich kein signifikanter Zusammenhang.
- * Von fünf geprüften Interaktionen sind zwei signifikant, eine ist tendenziell signifikant ausgeprägt.

Außerdem ist auf zwei in Abbildung 2.13. aufgeführte Arbeiten mit signifikanten Unterschieden hinzuweisen, die keine ökonomischen Effizienzgrößen verwenden. Nach Keen und Bronsema soll sich in der nicht beschaffbaren Arbeit von Ghani kein signifikanter Unterschied bei einem vergleichsweise

Abbildung 2.13.: Einflüsse Kognitiver Stile auf die Effizienz betrieblicher Entscheidungen (Teil 1)

- Impulsivität/Reflexivität (Matching Familiar Figures Test) -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Hussy/Gran-zow (Determinanten 1985)*	Mondlandung: Raumfähre zum Schwebestillstand bringen	160 Stud.	Impulsivität/Reflexivität, Kagans MFF-analoger Test lt. Wagner (Ablösung 1981); exp. Bed.: ohne/mit Rückmeldung, ohne/mit ext. Gedächtnis, transparent/intransparent (8 Versuchsgruppen)	Punktescore, berücks. Endgeschwindigkeit und Höhe	Reflexive sign. besser, (Extremgruppenvgl. für die erste Phase), Re-Analyse: punktbis. Korrr.: 0,11; Kein Interaktionstest
Henneman/Rouse (Fault Diagnosis Tasks 1984), 1. Exp.	Diagnose von Fehlern in fiktivem, "kontextfreiem" techn. System Netzwerkstellung	34 Stud.	Impulsivität/Reflexivität, Kagans MFF-Test a) Zeit bis erste Antwort b) Anzahl Fehler (beide Maße mit -0,61 korreliert)	Kosten ersetzter Komponenten ("COST")	a) Reflexive sign. besser Korrr.: - 0,38 b) Reflexive sign. besser Korrr.: 0,40
Henneman/Rouse (Fault Diagnosis Tasks 1984), 2. Exp.	Diagnose von Fehlern in kontextgebundenem System, "aircraft and automobile power plants"	26 Stud.	Impulsivität/Reflexivität, Kagans MFF-Test a) Zeit bis erste Antwort b) Anzahl Fehler (keine Angabe über Korrr. der Indikatoren)	Kosten ersetzter Komponenten ("COST")	a) n.s., Korrr.: -0,27 b) Reflexive sign. besser, Korrr.: 0,40

- Feldunabhängigkeit (Group Embedded Figures Test) -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Benbasat/Dexter (Value and Events 1979)	Unternehmensspiel, Beschaffung und Produktion	24 Stud. 20 Fak. Mitgl. 4 Manager	Feldunabhängigkeit, Group Embedded Figures Test, Medianteilung der Vpn; exp. Bed.: strukt., aggregierter Bericht vs. individuelle Abfragen im Dialog	Deckungsbeiträge	Feldunabhängige sign. besser, sign. Interaktion: bei Berichten Feldunabh. besser, bei indiv. Abfrage Feldabh.
Lusk (Differential Peaking 1979)*	Beantwortung von Faktenfragen	300 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, Medianteilung der Vpn; exp. Bed.: versch. Tabellen und Graphiken	Anzahl korrekt beantworteter Fragen	Feldunabhängige sign. besser, keine sign. Interaktion
Benbasat/Dexter/Masulis (Human Computer Interface 1981)	Finden einer optimalen Preis-Mengen-Kombination	50 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, Medianteilung der Vpn; exp. Bed.: a) Art der Dialogführung, b) Kommandos mit vs. ohne Abkürzung	a) Finden der optimalen Kombination b) Erzielter Gewinn	a) Feldunabhängige sign. häufiger b) n.s. Kein Interaktionstest
Benbasat/Dexter (Individual Differences 1982)	Unternehmensspiel, Beschaffung und Produktion	61 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, Medianteilung der Vpn; exp. Bed.: ohne vs. mit Entscheidungshilfe (Simulation)	Deckungsbeiträge	Feldunabh. besser, Haupteffekt tend. sign., kein Interaktionstest (ohne Hilfe Feldunabh. sign. besser, mit Hilfe beide gleich)

Abbildung 2.13.: Einflüsse Kognitiver Stile auf die Effizienz betrieblicher Entscheidungen (Teil 1, Fortsetzung)

- Feldunabhängigkeit (Group Embedded Figures Test), Fortsetzung -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Ghani/Lusk Information Representa- tion 1982)	Unternehmens- spiel, Absatz- mengenplanung	53 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, Me- dianteilung der Vpn; exp. Bed.: Tabelle vs. Graphik	Erzielte Gewin- ne ("profitabi- lity") und Ge- winnveränderg.	n.s. Kein Interaktions- test
Otley/Dias (Aggrega- tion 1982)	Unternehmens- spiel, Produkt- programmplanung	40 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, Me- dianteilung der Vpn; exp. Bed.: Aggregationsgrad	Schätzfehler 1. Kosten- prognose ("DIEE") 2. Gewinn- prognose ("CPEE")	1. Feldunabh. tend. sign. besser 2. n.s. Kein Interaktions- test
Zmud/Blo- cher/Mof- fie (Color Graphic 1983)	Bewertung von Risiken bei Rechnungsprü- fungen	51 In- terne Revi- soren	Feldunabhängigkeit, Group Embedded Figures Test; exp. Bed.: a) Tabellen vs. farbige Balkendiagramme b) 5 vs. 9 Informationen	Ob Risiko- klasse der Rechnung korrekt klassifi- ziert wird	n.s., durchschn. Korr.: 0,13; (Re- Analyse "Table 5" und "Table 6"), keine sign. Inter- aktion mit Präsen- tationsform tend. sign. Interaktion mit Angebotsmenge (Re-Analyse Tab. 5 u. 6 Vgl. d. Korr.)
Henneman/ Rouse (Fault Diagnosis Tasks 1984), 1. Exp.	Diagnose von Fehlern in fiktivem, "kon- textfreiem" techn. System in Netzwerk- darstellung	34 Stud.	Feldunabhängigkeit Embed- ded Figures Test: a) Lösungszeiten b) Anzahl Fehler (beide Maße mit 0,57 kor.)	Kosten er- setzter Kom- ponenten, ("COST")	a) tend. sign. Korr.: 0,28 b) n.s. Korr.: 0,19
Henneman/ Rouse (Fault Diagnosis Tasks 1984), 2. Exp.	Diagnose von Fehlern in kontextgebun- denem System, "aircraft and automobile power plants"	26 Stud.	Feldunabhängigkeit, Embed- ded Figures Test: a) Lösungszeiten b) Anzahl Fehler (keine Angabe über Korre- lationen bei diesem Exp.)	Kosten er- setzter Kom- ponenten, ("COST")	a) n.s. Korr.: -0,11 b) n.s. Korr.: 0,08
Benbasat/ Dexter/ (Graphical Information Presenta- tion 1985)	simul. Verkaufs- einsatzplanung, Aufteilung des Budgets auf ver- schiedene Ver- kaufsgebiete	35 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, Me- dianteilung der Vpn; exp. Bed.: a) Tabelle vs. Gra- phik, b) ohne vs. mit Farbe	Erzielte Gewin- ne, relativ zum Optimum	1. Feldunabh. sign. besser 2. Interaktion mit Präsentations- form n.s. 3. Interaktion mit Farbe sign.
Fink (Ko- gnitive Stile 1985)	Bilanzanalyse für Kreditent- scheidung	36 Stud.	Feldunabhängigkeit, Group Embedded Figures Test, als kontinuierliche Skala ver- wendet; getrennt nach: 1. novativer Fall 2. Wiederholungsfälle Rangkorrelationen und quadr. Regressionen	Treffsicher- heit der Ana- lyse (Gesamt- lage)	1. n.s. Korr.: -0,06 2. tend. sign. (bei zweiseit. Test), Korr.: 0,26 Keine sign. nicht- linearen Zusammen- hänge

*: Studie wird nur nachrichtlich vermerkt, da kein ökonomisches Effizienzmaß

komplexen, computergestützten Unternehmensspiel ergeben haben.³¹⁾

Die positive Bilanz ist mit der Entsprechung der kognitiven Fähigkeiten, die die Tests erfassen, mit den Anforderungen der Aufgaben, an denen die Effizienzprüfung vorgenommen wird, zu erklären:³²⁾

1. Der "Group Embedded Figures Test" ist ein graphischer Perzeptionstest, bei dem einfache geometrische Figuren in unterschiedlichen graphischen Kontexten wiederzuerkennen sind. Beim "Matching Familiar Figures Test" werden ein Originalbild (z. B. das Bild einer Rose) und mehrere abgewandelte Bilder vorgelegt. Aufgabe ist die Identifikation der einzigen Kopie, die dem Original genau entspricht. Beziehungsanalysen mit anderen Tests deuten darauf hin, daß mit diesen Tests nicht nur elementare Wahrnehmungsvorgänge, sondern auch in gewissem Umfang räumliches Vorstellungsvermögen und quantitative Fähigkeiten erfaßt werden.³³⁾

2. Die Aufgaben, in denen die perzeptionsbezogenen Stile Unterschiede erklären, sind vergleichsweise einfacher Natur. Sie besitzen im großen und ganzen gesehen folgende Merkmale:

- * Es existiert eine monovariable Zielfunktion mit einer eindeutigen Lösung, die sich als "richtig" oder "falsch" bzw. "optimal" oder "nicht optimal" klassifizieren läßt.
- * Das Informationsangebot ist vollständig und überschaubar: Alle zur Lösung erforderlichen Informationen werden in übersichtlicher Form angeboten. Es müssen keine Informationen aktiv gesucht werden.
- * Zur Lösung des Problems müssen die angebotenen Informationen entweder aus einem graphischen Kontext gelöst und restrukturiert werden, oder es sind elementare mathematische Operationen zu vollziehen.

Dieses scheinen Problemanforderungen zu sein, die den gemessenen Fähigkeitsunterschieden entgegenkommen. Die geforderte **Entsprechung von Persönlichkeit und Aufgabe** scheint erfüllt.

Die signifikanten Interaktionseffekte und die nichtsignifikanten Ergebnisse der Untersuchungen zum "Group Embedded Figures Test" weisen jedoch auf die **Grenzen der Effizienzwirkungen** der gemessenen Fähigkeiten hin:

- * Nur bei einer fiktiven, "kontextfreien", graphisch repräsentierten technischen Anlage sind Feldunabhängige überlegen. Bei der Suche von Fehlern, die in einer realen Anlage auftreten, verschwindet die Überlegenheit. Sie kann offenbar durch konkrete Erfahrung der Diagnostiker mit diesem System substituiert werden.³⁴⁾
- * Nur bei der einfachen Beurteilungssituation mit fünf Informationen sind Feldunabhängige etwas überlegen, bei neun Informationen ergeben sich keine signifikanten Unterschiede.³⁵⁾
- * Nur bei strukturierten Berichten, die alle problemrelevanten Informationen enthalten, sind Feldunabhängige überlegen. Bietet man den Feldabhängigen ein einfaches Simulationsmodell zum Durchspielen der Auswirkungen ihrer Entscheidungen an, wird dieser Nachteil ausgeglichen.³⁶⁾ Bei individueller Datenabfrage von einer Datenbank sind sie Feldunabhängigen sogar überlegen.³⁷⁾ Hier scheinen zwei Einflüsse zusammenzukommen: Erstens können Feldunabhängige nicht ihren Vorteil ausspielen, ein gegebenes, problemgerechtes Informationsangebot besser zu strukturieren, zweitens neigen Feldabhängige offenbar eher dazu, explorativ vorzugehen und Informationen aktiv zu suchen.³⁸⁾

Fazit: Zwischen feldunabhängigen und feldabhängigen Personen lassen sich konsistent signifikante Leistungsunterschiede feststellen, die mit dem in der psychologischen Forschung gut belegten Konstrukt erklärt werden können. Die Leistungsunterschiede gelten jedoch nur für gut-strukturierte Aufgaben vergleichsweise geringer Komplexität: Falls ein semantischer Kontext oder Entscheidungshilfen vorhanden sind, oder falls die Informationen aktiv gesucht werden müssen, verschwindet der Vorteil Feldunabhängiger. Er kann sich im letzten Fall sogar in einen Nachteil umkehren.

2.3.3.2. Myers-Briggs Indikatoren: Keine praktisch bedeutsamen Einflüsse eines besonders häufig empfohlenen Stilkonzeptes

Die empirische Evidenz für die besonders häufig und prononciert empfohlenen Stilmaße von Myers-Briggs, die auf die Persönlichkeitstheorie von Jung zurückgehen, ist besonders ungünstig ausgeprägt (vgl. Abbildung 2.14.):³⁹⁾

- * Nur zwei der vierzehn getesteten Haupteffekte sind signifikant ausgeprägt. Die beiden signifikanten Ergebnisse widersprechen sich.⁴⁰⁾ Zwei der vierzehn Tests zeigen einen tendenziell signifikanten Einfluß, der jedoch den Erwartungen der Autoren widerspricht.⁴¹⁾

- * Keiner der fünf getesteten Interaktionseffekte ist signifikant.

Hinzuzufügen ist, daß die Effektstärken im allgemeinen sehr gering sind und die Stichproben über 50 liegen, so daß die hypothesenkonträre Evidenz nicht auf mangelnde Teststärke zurückgeführt werden kann.

Das Ergebnis steht sowohl zu den Befunden der perzeptionsbezogenen Stilmaße als auch zur positiven Bestandsaufnahme von Keen und Bronsema in deutlichem Widerspruch. Daher sind einige differenzierende Überlegungen erforderlich.

1. Die Abweichung gegenüber den perzeptionsbezogenen Stilmaßen kann sowohl auf unterschiedlichen Anforderungen der untersuchten Aufgaben als auch auf unterschiedlichen Inhalten der gemessenen Persönlichkeitsvariablen beruhen. Da in den Effizienzprüfungen beider Stilmaße vergleichbare Aufgaben untersucht werden, ist die Ursache eher bei den Stilmaßen zu suchen. Im Unterschied zu den perzeptionsbezogenen Stilmaßen erfassen die Myers-Briggs-Indikatoren keine spezielle Fähigkeit, sondern eine generelle **Neigung**, Probleme in einer bestimmten Weise anzupacken. Die Antworten auf die hierzu gestellten Fragen sind nicht mit "ja" oder "nein" zu beantworten. Sie sollen vielmehr dazu beitragen, "bevorzugte Wahrnehmungs- und Entscheidungsmodi aufzuzeigen".⁴²⁾

Abbildung 2.14.: Einflüsse Kognitiver Stile auf die Effizienz betrieblicher Entscheidungen (Teil 2)

- Jungsche Persönlichkeitstypen (Myers-Briggs-Type-Indicators) -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Baxendale (Decision Accuracy 1978)	Analyse von Kostenabweichungen	99 Stud.	Myers-Briggs-Type-Dimension Denken/Fühlen; exp. Bed.: numerische vs. verbale Informationen	Anzahl korrekt klassifizierter Abweichungen	n.s., erkl. Varianz 0,1 v.H. (Re-Analyse "Table 5"), keine sign. Interaktion
Casey (Usefulness 1980)	Beurteilung, ob ein Unternehmen in den nächsten drei Jahren insolvent wird	46 Kreditmanager	Myers-Briggs-Type-Dimension Sinneswahrnehmung/Intuition, Bildung von zwei Gruppen, 67 v. H. "sensors"	Anzahl korrekt klassifizierter Firmen	"sensors" sign. schlechter, Korrr.: -0,30
Davis/El-nicki (Cognitive Types 1984) vgl. auch Davis (Information Presentation 1980)	Unternehmensspiel, Abstimmung von Produktions- und Absatzmengen	96 Stud.	Myers-Briggs-Test-Maße: a) Sinnesw. / Intuition b) Denken/Fühlen Bildung von je zwei Gruppen führt zu vier Typen; exp. Bed.: a) Tabelle vs. Graphik, b) unverdichtete vs. verdichtete Daten	Kosten	a) "sensors" sign. besser, erkl. Var.: 5,8 v.H. b) "feeling" tend. sign. besser, erkl. Varianz: 2,8 v.H. (Re-Analysen von "Table 4") Keine Interaktionstests, nach Typen diff. Auswertung zeigt Unterschiede
McIntyre (Marketing Models 1982)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verkaufsgebiete	69 Stud.	Myers-Briggs-Test-Maße: a) Extra/Introversion b) Sinneswahrng./Intuition c) Denken/Fühlen d) Urteilen/Wahrnehmen exp. Bed.: Entscheidungshilfe, Verkaufsgebiete, Zufallsschwankungen	Erzielte Gewinne, relativ zum Optimum	a) n.s. b) n.s. c) n.s. d) n.s. keine sign. Interaktionen mit Entscheidungshilfe, and. Bedingungen, kein Test
Ghani/Lusk (Information Representation 1982)	einfaches Unternehmensspiel, Absatzmengenentscheidung unter Unsicherheit	53 Stud.	Myers-Briggs-Test-Maße: a) Sinneswahrng./Intuition b) Denken/Fühlen Bildung von 4 Typen aus beiden Dimensionen; exp. Bed.: Tabelle vs. Graphik und Konstanz/Wechsel der Präsentationsform	Erzielte Gewinne ("profitability") und Gewinnveränderg.	a) n.s. b) n.s. Kein Interaktionstest
Fink (Kognitive Stile 1985)	Bilanzanalyse für Kreditentscheidung	36 Stud.	Myers-Briggs-Test-Maße: a) Extra/Introversion b) Sinneswahrng./Intuition c) Denken/Fühlen d) Urteilen/Wahrnehmen getrennt nach: 1. novativer Fall 2. Wiederholungsfälle Rangkorrelationen und quadr. Regressionen	Treffsicherheit der Analyse (Gesamt-lage)	a) 1. n.s. Korrr.: -0,04 2. n.s. Korrr.: -0,03 b) 1. n.s. Korrr.: 0,01 2. tend. sign. Korrr.: 0,22 c) 1. n.s. Korrr.: -0,08 2. n.s. Korrr.: 0,18 d) 1. n.s. Korrr.: -0,18 2. n.s. Korrr.: -0,08 Keine sign. nicht-linearen Zusammenh.

Dies geschieht beispielsweise durch Stellen folgender Fragen:⁴³⁾

- * "Hören Sie öfter (A) auf ihr Herz oder (B) auf Ihren Verstand?"
- * "Schätzen Sie normalerweise (A) Gefühle mehr als Logik oder (B) Logik mehr als Gefühle?"
- * "(A) Ist es leicht oder (B) ist es schwierig Sie kennenzulernen?"
- * "Welches Wort in jedem Paar ist für sie ansprechender? (...) (A) weich - hart (B); (A) wer - was (B); (A) konkret - abstrakt (B)".

Gemessen an diesen sehr allgemein gehaltenen Selbsteinschätzungen mutet es überraschend an, daß man ohne Bezug zu den Anforderungen konkreter Aufgaben mit starken Interaktionseffekten rechnet.⁴⁴⁾ Gegen solche Erwartungen sprechen folgende Überlegungen: Die Stilmaße erfassen eine Präferenz für eine wahrgenommene **Verhaltensintention**, die unter "normalen" Bedingungen gelten soll und nicht das **tatsächliche** Verhalten bei einer konkreten Aufgabe. Man darf davon ausgehen, daß das faktische Verhalten eine beachtliche Bandbreite aufweist, deren Ausschöpfung stark von Situation und Problem abhängt. Um die Hypothese zu **prüfen**, daß sich bestimmte Typen auch bei unterschiedlichen Aufgaben anders verhalten als andere und deshalb Leistungsunterschiede auftreten, hätte man dieselben Personen mehrere unterschiedliche Aufgaben lösen lassen müssen. Dies ist unseres Wissens nicht geschehen. Nur Fink prüft, ob sich bei novativen Fällen andere Zusammenhänge ergeben als bei Wiederholungsfällen (allerdings bei Konstanthaltung des Problems "Bilanzanalyse"). Sie tun es nicht.⁴⁵⁾ Wie unsere Bestandsaufnahme zeigt, ergeben sich auch zu Parametern des Informationssystems keine signifikanten Interaktionen und gerade darauf zielte der stark beachtete konzeptionelle Beitrag von Mason und Mitroff.

2. Keen und Bronsema referieren zahlreiche Studien, die belegen, daß Mitglieder unterschiedlicher Bevölkerungs- und Berufsgruppen unterschiedlich ausgeprägte Stilmeßwerte aufweisen.⁴⁶⁾ Man kann mit diesen Untersuchungen jedoch nicht beweisen, daß erfolgreiche Manager andere Stilmerkmale aufweisen als nicht erfolgreiche, oder etwa, daß Kreditmanager der zweiten hier-

archischen Ebene ein bestimmtes Stilprofil benötigen, um ihre Aufgabe besonders gut zu erfüllen. Es erscheint durchaus plausibel, daß man wegen unterschiedlicher Aufgaben, die man "normalerweise" löst, auch eine andere Selbsteinschätzung des Kognitiven Stiles gewinnt. Dies bedeutet aber nicht, daß man bei einem konkreten Problem, welches andere als die üblichen Anforderungen stellt, auch eine andere Verhaltensweise als die "normalerweise" präferierte wählen wird. Außerdem zeigen die Befunde bei Keen und Bronsema beachtliche Bandbreiten innerhalb der verglichenen Berufsgruppen, so daß man aus bestimmten Stilwerten keinesfalls auf **erfolgreiche** Berufsausübung schließen sollte.

Fazit: Die von Jung konzipierten Persönlichkeitsdimensionen eignen sich nicht für eine Effizienzprognose in der häufig verwendeten Version der Myers-Briggs-Indikatoren, auch nicht für eine differentielle Erklärung von Effizienzunterschieden, die auf Interaktionseffekte abstellt. Man kann dafür weder die mangelnde Reliabilität der Meßinstrumente verantwortlich machen, denn sie ist hoch, noch die Wissenschaftler, die diese Instrumente entwickelt haben, denn sie hatten andere Zwecke im Auge. Verantwortlich für die geringe Validität des Instrumentes bezüglich der Effizienz der untersuchten Aufgaben sind die Anwender, die den Test benutzt haben. Wir wenden uns nicht prinzipiell dagegen, psychometrisch geeichte Instrumente auch für andere als die ursprünglich vorgesehenen Zwecke einzusetzen und zu versuchen, ihren Geltungsbereich explorativ zu vergrößern. Man sollte dieses nur deutlicher hervorheben und aus der Tatsache, daß man keine beachtlichen Zusammenhänge feststellt, angemessene Konsequenzen ziehen.

2.3.3.3 "Analytisch" vs. "Heuristisch": keine praktisch bedeutsamen Einflüsse selbst-gestrickter Meßkonzepte

Zeigen die von amerikanischen Betriebswirten entwickelten Stilmaße bessere Ergebnisse? Abbildung 2.16. beantwortet diese Frage mit einem klaren Nein:

* Nur in einer der acht Studien ergibt sich ein signifikanter Haupteffekt.

Abbildung 2.15.: Einflüsse Kognitiver Stile auf die Effizienz betrieblicher Entscheidungen (Teil 3)

- "Analytisch-Heuristischer Stil" (versch. Meßinstrumente) -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Mock/ Estrin/ Vasar- helyi (Learning 1972)	Unternehmens- spiel, Pro- duktion und Absatz	31 Man. und Stud.	Analytisch-Heuristisch, eigener Fragebogen ("Ap- pendix C"), drei Gruppen: Heuristiker, Analytiker, Sonstige; exp. Bed.: un- mittelbarer vs. verzöger- ter Feedback	1. Gewinne 2. Kosten	1. Analytiker sign. höhere Gewinne, erkl. Varianz: 11,6 v. H. 2. Analytiker sign. geringere Kosten erkl. Varianz: 35,8 v. H. (Re-Analyse von "Table 4") Kein Interaktions- test
Benbasat (Effects 1974)	Unternehmens- spiel, Beschaf- fung und Pro- duktion	32 Stud.	Analytisch-Heuristisch, Fra- gebogen nach Barkin, basiert z. T. auf Myers-Briggs-Test, Medianteilung der Vpn; exp. Bed.: Angebotsmenge, Präsen- tationsform, Entscheidungs- hilfen, "exception reporting"	Kosten	n.s., keine sign. Interaktionen mit Wissens-Test und exp. Bedingungen
Vasarhelyi (Planning Systems 1977)	simul. Finanz- planungspro- blem	50 Stud. und Mana- ger	Analytisch-Heuristisch: Fra- gebogen, ohne nähere Erläu- terung, Verweis auf unveröff. Diss. u. ältere Fassung bei Mock/Estrin/Vasarhelyi	Experten- urteil zur Qualität der Pla- nung	n.s. Kein Interaktions- test
Lucas (Graphics 1981)	Unternehmens- spiel, Bestell- mengenentschei- dung unter Un- sicherheit	119 Mana- ger	Analytisch-Heuristisch, Fragebogen nach Barkin; exp. Bed.: nur tabella- rich vs. nur graphische Darst. vs. tabellarische und graphische Darstellg.	Kosten ("best si- mulation")	n.s.; tend. sign. Interaktion: mit Graphik "Heuristi- ker" besser, bei "Analytiker" kein Einfluß der Prä- sentationsform
Mock/Vasar- helyi (Co- gnitive Style 1984) Experiment IS4	Unternehmens- spiel, Produk- tion und Ab- satz	130 Stud.	Analytisch-Heuristisch: Fra- gebogen, ohne nähere Erläu- terung, Verweis auf nicht im Literaturverzeichnis auf- geführte Studie; exp. Bed.: a) Vollständigkeit, b) Fein- heit ("coarseness") des Feed- back über Budgetabweichungen	Gewinne ("average profits")	n.s., keine sign. Interaktionen; keine Effekt- schätzung wg. mangelnder An- gaben
Mock/Vasar- helyi (Co- gnitive Style 1984) Exp. SX1	Börsenspiel mit 50 Papieren der New Yorker Börse	97 Stud.	Analytisch-Heuristisch: Fra- gebogen, ohne nähere Erläu- terung, Verweis auf nicht im Literaturverzeichnis auf- geführte Studie	Erzielte Rendite	n.s., kein Inter- aktionstest; keine Effektschätzung wg. mangelnder An- gaben
Fink (Ko- gnitive Stile 1985)	Bilanzanalyse für Kreditent- scheidung	36 Stud.	Analytisch-Heuristisch, Fra- gebogen nach Barkin, basiert z. T. auf Myers-Briggs-Test, dt. Fassung von Fink, Befun- de getrennt nach: 1. novativer Fall 2. Wiederholungsfälle Rangkorrelationen und quadr. Regressionen	Treffer- sicherheit der Analyse (Gesamtlage)	1. n.s. Korr. -0,15 2. n.s. Korr. -0,13 Keine sign. nicht- linearen Zusammenh.

- * Von acht geprüften Interaktionseffekten ist nur einer tendenziell signifikant.

Ein Teil der Studien verwendet den von Barkin "entwickelten" Fragebogen. Er ist hoch korreliert mit der Myers-Briggs-Dimension "Judging/Perceiving", weil seine Fragen zum Teil wörtlich aus diesem Test übernommen werden.⁴⁷⁾ Daher gelten für ihn unsere im letzten Kapitel gemachten Aussagen. Von Barkin zusätzlich vorgeschlagene Likertskalen-Items wie "Ich bin am besten, wenn ich einem Plan folge"⁴⁸⁾, sind wenig geeignet, unsere grundsätzliche Kritik zu entkräften.

Ist das von Mock, Estrin und Vasarhelyi verwendete Maß besonders geeignet für eine Effizienzprognose, weil es als einziges einen signifikanten Unterschied zwischen "Analytikern" und "Heuristikern" anzeigt? Nein, denn gegen deren Messung sind erhebliche Bedenken zu erheben. Die Autoren nehmen eine Klassifikation der Versuchspersonen nach deren **nachträglich** erfragter Verhaltensbeschreibung vor. Nach welchen Kriterien die Zuordnung der Antworten auf die offene Frage "How did you about reaching the various decisions for this experiment" erfolgt, wird nicht expliziert.⁴⁹⁾ Man erfährt auch nichts über die Reliabilität der Messung oder die Effizienzmeßwerte der nicht klassifizierten Personen. Daher liegt keine reproduzierbare Messung vor.

Fazit: Die von amerikanischen Betriebswirten entwickelten Maße zu der schillernden Stildimension "Analytisch/Heuristisch" liefern keinen beachtlichen Beitrag zur Effizienzprognose. Sie sind hinsichtlich Reproduzierbarkeit, Reliabilität und Validität der Messung deutlich schlechter einzuschätzen als die in der psychologischen Forschung gut belegten "Myers-Briggs-Type-Indicators". Der hier begangene Pfad ist aus heutiger Sicht als Irrweg anzusehen.⁵⁰⁾

2.3.3.4. Sonstige Stile: Tendenzielle Einflüsse, die weiterer Prüfung bedürfen

Die Restmenge der "sonstigen Stile", die einer Effizienzanalyse unterzogen werden, ist recht heterogen. Daher sind kaum verallgemeinerungsfähige Aus-

sagen zu treffen (vgl. Abbildung 2.16.):

1. Daß sich in der Untersuchung von Craft und Koautoren kein signifikanter Zusammenhang ergibt, dürfte an der Art der Messung von **Rowe's decision style** liegen: Likertskalen-Items wie "I dislike losing control, boring work, following rules, being rejected" können unseres Erachtens nur eine sehr unspezifische Verhaltenspräferenz erfassen.

2. Die drei Studien mit **Driver's "decision style"**, der von den Autoren als Weiterentwicklung der eindimensionalen Typologie von Schroder, Driver und Streufert angesehen wird, zeigen keine signifikanten Unterschiede. Da zum Meßverfahren nur spärliche Aussagen gemacht werden, kann über die Ursachen der fehlenden Zusammenhänge nur spekuliert werden. Anzumerken ist, daß weder Savich noch McGhee/Shields/Birnberg signifikante Einflüsse auf das Informationsverhalten feststellen konnten.⁵¹⁾

3. Zur **relativen Hemisphärendominanz** ist uns nur die Effizienzprüfung von Fink bekannt. Der relativ stark ausgeprägte umgekehrt-u-förmige Zusammenhang entspricht seiner Erwartung. Dies könnte ein Anlaß sein, Effizienzwirkungen dieses Konzeptes in zukünftigen Untersuchungen erneut zu testen.

4. Zu dem insbesondere in der verhaltensorientierten Rechnungswesenforschung recht stark beachteten Konstrukt der **Kognitiven Komplexität** liegen überraschend wenige Effizienzprüfungen vor, da in der Rechnungswesenforschung in der Regel das Informationsverhalten zur Kognitiven Komplexität in Beziehung gesetzt wird. Die empirische Evidenz ist kontrovers ausgeprägt: Während Pratt keine signifikanten Effekte ermittelt, stellt Fink einen recht starken signifikanten Zusammenhang mit dem Effizienzmaß "Treffsicherheit der Beurteilung" fest. Dies gilt jedoch nur für die Operationalisierungsvariante "Diskriminierung". Für die weniger reliabel gemessene "Differenzierung" zeigt sich kein signifikanter Zusammenhang. Bei Pratt ergibt sich für die ex-post gemessene Kognitive Komplexität der Teilnehmer bei der Versuchsbedingung mit einem besonders komplexen Jahresabschluß ebenfalls eine signifikante Beziehung. Diese Versuchsbedingung dürfte der von Fink untersuchten am ehesten entsprechen, so daß sich der Widerspruch auflösen würde. Für diese Erklärung würde auch der bei Malhotra festgestellte signi-

Abbildung 2.16.: Einflüsse Kognitiver Stile auf die Effizienz betrieblicher Entscheidungen (Teil 4)

- "Kognitive Komplexität" (versch. Meßinstrumente) -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Malhotra (Information Load 1982)*	Beschaffung von Häusern	300 Haushaltsvorstände	Kognitive Komplexität, Pinsons Variante des "retests" von Biery, exp. Bed.: Anzahl angebotener Alternativen und Attribute	Ob die jeweils nutzenmax. Alternative gewählt wird	Kogn. Komplexe sign. besser (Kovariate in Logit-Analyse), kein Interaktionstest
Pratt (Post-Cognitive Structure 1982)	Analyse von Jahresabschlüssen	90 Stud.	Kognitive Komplexität, MDS-Skalierung der wahrgenommenen Ähnlichkeit von 12 Teilen des Reports: a) vor Analyse des Jahresabschlusses b) nach Analyse und mit Bezug auf exp. Jahresabschluß Exp. Bed.: Komplexität der Jahresabschlüsse	Güte der Prognose des Netto-Erfolges	a) n.s., Korr. 0,09 b) n.s., Korr. 0,11 Keine sign. Interaktion (Re-Analyse von Tab. 4 u. 6 mittels Fisher's z-Transf., für Interaktionstest Unterschiede der Korr. auf Signifikanz getestet)
Fink (Kognitive Stile 1985)	Bilanzanalyse für Kreditentscheidung	36 Stud.	Kognitive Komplexität, REP-Test, dt. Fassung von Bronner (Komplexität 1983), zwei Stildimensionen: a) Differenzierung b) Diskrimination (korrelieren mit 0,45); Befunde getrennt nach: 1. novativer Fall 2. Wiederholungsfälle Rangkorrelationen und quadr. Regressionen	Treffer-sicherheit der Analyse (Gesamt-lage)	a) 1. n.s. Korr.: -0,09 2. n.s. Korr.: -0,12 b) 1. diskrim. Pers. sign. besser Korr.: -0,55 2. diskrim. Pers. sign. besser Korr.: -0,31 Keine sign. nicht-linearen Zusammenh.

- Rowe's "Decision Style" -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Craft/Mock/Rowe/Ryan ("Decisive" Style 1984)	Unternehmensspiel, Produktion und Absatz	30 Stud.	Rowe's Decision Style Inventory, Bildung von 4 Typen, Messung im Anhang beschrieben; exp. Bed.: 2x2-Design Feedback und Komplexität der Berichte, nur Befunde für "komplexe Berichte" ausgewiesen ("Exhibit 4", S. 12)	1. Gewinne ("average income") 2. Prognosegüte	1. n.s., erkl. Var.: 6,3 v. H. 2. n.s., erkl. Var.: 9,3 v. H. Keine Interaktionstests, Re-Analyse v. Exhibit 4 zur Best. der erkl. Varianz

(Anmerkung: Die Hypothesen der Autoren beziehen sich ausschließlich auf "komplexe Berichte". Daher werden nur für diese Versuchsbedingung Befunde ausgewiesen. Aus den etwas unklaren Angaben im Text S. 10 kann man schließen, daß sich für das gesamte Sample von N=57 Studenten ebenfalls keine sign. Unterschiede ergeben. Die erklärte Varianz müßte bei den "einfachen" Berichten noch geringer ausfallen, denn wg. der größeren Stichprobe müßte sonst ein signifikanter Haupteffekt auftreten).

Abbildung 2.16.: Einflüsse Kognitiver Stile auf die Effizienz betrieblicher Entscheidungen (Teil 4)

- Driver's "Decision Style" -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Driver/Mock (Experimental Results 1976)	Unternehmensspiel, Produktion und Absatz	54 Stud.	"Decision-Style": Verweis auf Driver/Mock 1975, wo dessen Messung grob beschrieben wird, Fragebögen nicht zugänglich, exp. Bed.: Vollständigkeit des Feedback über Budgetabweichungen	Gewinne ("average profits")	n.s. Kein Interaktionstest wg. ungleicher Zellenbesetzungen, Einzeltests bei untersch. Angebotsbed. deuten auf Interaktion hin
Mock/Vasarhelyi (Cognitive Style 1984) Experiment IS4	Unternehmensspiel, Produktion und Absatz	130 Stud.	"Decision-Style": Verweis auf Driver/Mock 1975, wo dessen Messung grob beschrieben wird, Fragebögen nicht zugänglich, exp. Bed.: a) Vollständigkeit, b) Feinheit des Feedback über Budgetabweichg.	Gewinne ("average profits")	n.s., keine sign. Interaktionseffekte
Mock/Vasarhelyi (Cognitive Style 1984) Exp. SX1	Börsenspiel mit 50 Papieren der New Yorker Börse	97 Stud.	"Decision-Style": Verweis auf Driver/Mock 1975, wo dessen Messung grob beschrieben wird, Fragebögen nicht zugänglich	Erzielte Rendite	n.s., kein Interaktionstest

- Hemisphärendominanz (Herrmann Dominance Test) -

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Fink (Kognitive Stile 1985)	Bilanzanalyse für Kreditentscheidung	36 Stud.	Herrmann Dominance Test, Skala verbal/analytischer Stil, Skalierungsverfahren nicht veröff.; diff. nach: 1. novativer Fall 2. Wiederholungsfälle Nur Ergebnisse von quadr. Regressionen berichtet.	Treffsicherheit der Analyse (Gesamt-lage)	1. n.s. 2. sign. nicht-linearer Zusammenhang, erklärte Varianz: 20 v. H.

*: Studie wird nur nachrichtlich vermerkt, da kein ökonomisches Effizienzmaß

fikante Einfluß sprechen, denn in seinem Experiment werden besonders große Attribut- und Alternativenmengen zur Beurteilung vorgelegt. Bei dieser positiven Einschätzung der Befunde sollte jedoch nicht übersehen werden, daß nicht nur in diesen drei Studien, sondern generell außerordentlich divergierende, empirisch nur schwach korrelierende Instrumente zur Messung des Konstruktes "Kognitive Komplexität" verwendet werden. Man muß daher auch bei diesem Stil abwarten, ob weitere Studien mit vergleichbaren Aufgabenstellungen und Stilmaßen konsistente Ergebnisse zeigen.

Fazit: Die Untersuchungen zur heterogenen Restmenge lassen nur den Schluß zu, daß Driver's "decision style" wenig geeignet ist, Effizienzunterschiede zu erklären. Bei den Konstrukten "Kognitive Komplexität" und "Hemisphären-dominanz" sind die Ergebnisse weiterer Effizienzprüfungen abzuwarten.

2.3.4. Einflüsse von Test-Intelligenz und quantitativen Fähigkeiten

2.3.4.1. Test-Intelligenz: Effizienzeinflüsse bei test-gerechten Aufgaben

Die Beziehungen zwischen Intelligenz, Denken und Problemlösen werden in der psychologischen Forschung schon seit langem erforscht. Erst in jüngerer Zeit wird jedoch untersucht, ob die in gängigen Tests gemessene Test-Intelligenz ein Prädiktor für das erfolgreiche Lösen komplexer, eigendynamischer, intransparenter, vernetzter Probleme ist, die man mit Computersimulationen erzeugt. Die untersuchten Probleme decken eine recht große Bandbreite ab: Sie reichen von vergleichsweise überschaubaren "Mondlande"- und "Kochproblemen", bei denen nur wenige Aktionsparameter zu beeinflussen sind, bis hin zu hochkomplexen Simulationen, in denen man eine Vielzahl der wirtschaftlichen und politischen Entscheidungen einer Gemeinde treffen kann.⁵²⁾ Die Ergebnisse dieser Forschung werden in der psychologischen Forschung intensiv und kontrovers diskutiert.⁵³⁾ Da viele Studien ökonomische Effizienzmaße verwenden und betriebswirtschaftliche Aktionsparameter beeinflußt werden können, ist diese Teilmenge auch für unsere Untersuchung von erheblicher Bedeutung.⁵⁴⁾ Wir stellen deshalb auch hier die Frage, ob das untersuchte Persönlichkeitsmerkmal geeignet ist, ökonomische Effizienz-

variablen zu prognostizieren, sei es als Haupteffekt oder in Interaktion mit Problemstellung und Informationsangebotssituation.

Abbildung 2.17. belegt das Ergebnis unserer Bestandsaufnahme:

- * Von 15 getesteten Beziehungen ist nur eine signifikant ausgeprägt, eine Studie zeigt einen tendenziell signifikanten Einfluß.⁵⁵⁾

- * Von drei Interaktionstests ist einer signifikant.

Das Ergebnis fällt überraschend schlecht aus. Es erklärt die intensive Diskussion in der psychologischen Forschung, in der man offenkundig mit wesentlich engeren Zusammenhängen von Test-Intelligenz und Problemlöseerfolg gerechnet hatte bzw. noch immer rechnet.

Jene deutschen Psychologen, die für ihre Studien komplexe, computersimierte Managementprobleme wählten, führen dieses Scheitern vor allem auf die sehr unterschiedlichen Anforderungen von Intelligenz-Tests und komplexen Problemen zurück. Kreuzig, ein Mitarbeiter des viel beachteten "Lohhausen"-Projektes, schreibt hierzu:

"Um im Lohhausen-Versuch erfolgreich abzuschneiden, muß man ... die Vielfalt der verfügbaren Daten - durch Ordnen strukturiert und reduziert - im Gedächtnis verfügbar halten. Man muß die vorhandenen Problemfelder aktiv suchen, in ihrer Bedeutung erkennen und sich ein zutreffendes Bild von ihnen verschaffen. Man muß sich über die angestrebten Ziele klar werden und nach richtig dosierten, geeigneten Maßnahmen suchen, um sie zu verwirklichen. Man muß planen und entscheiden, dann aber auch tatsächlich handeln. (...) Man muß eventuelle Nebenwirkungen der beschlossenen Maßnahmen analysieren, die Effekte der Eingriffe kontrollieren und gegebenenfalls auch korrigieren. Und dies alles muß im Rahmen des zur Verfügung stehenden Zeit-Budgets erfolgen. Darüber hinaus muß man auch damit zurecht kommen, daß sich alles verändert, selbst, wenn man nicht eingreift."⁵⁶⁾

Abbildung 2.17.: Einflüsse der Testintelligenz auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Dentler (Problem-löseverhalten 1977)	agrарwirtschaftlich-ökologisches Managementspiel	30 Stud.	LPS-Test von Horn ("Leistungs-Prüf-System"), mittlere Intelligenzquotienten der 10 Gruppen (Tab. 59)	Rangsumme von sieben Effizienzmaßen (Tab 48-54)	n.s., Spearman'sche Rangkor.: -0,01 (Re-Analyse auf Basis von Tab. 48-54, 59)
Putz-Osterloh (Problem-löseerfolg 1981)	Unternehmensspiel, Produktion und Absatz	48 Stud.	APM-Test von Raven; exp. Bed.: ohne vs. mit Transparenz über Variablenzustände und Beziehungen	Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	n.s., kein Interaktionstest, (Rangkor. transp Bed.: n.s., intransp. Bed.: -0,22; kein Test auf Unterschiedlichkeit wg. mang. Inf.)
Putz-Osterloh/Lüer (Problem-löseleistungen 1981)	Unternehmensspiel, Produktion und Absatz	68 Stud.	APM-Test von Raven; exp. Bed.: ohne vs. mit Transparenz über Variablenzustände und Beziehungen	Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	n.s., kein Interaktionstest, (Rangkor. transp Bed.: 0,31 intransp. Bed.: n.s.; kein Test auf Unterschiedlichkeit wg. mang. Inf.)
Hesse (Semantischer Kontext 1982)	agrарwirtschaftlich-ökologisches Managementspiel bzw. rein formales System	120 Stud.	APM-Test von Raven; exp. Bedingungen: 1. ohne/mit Pfeildiagramm der Variablenbeziehungen, 2. ohne/mit semantischem Kontext	aufsumm. z-Wert über versch. Gütekriterien	sign. pos. Kor.: 0,25; sign. Interaktion mit semant. Kontext, Interaktion mit Pfeildiagramm n.s. (Re-Analyse mit Fischers z-Transf.)
Stäudel (Problemlösen 1982)	Managementspiel, ökologische und wirtschaftliche Entscheidungen	43 Stud.	CFT3 von Catell	6 versch. Effizienzmaße	n.s., kein Interaktionstest, Kor. bei Kapital -0,20, sonst zw. + 0,10
Funke (Testintelligenz 1983)	Unternehmensspiel, Produktion und Absatz	53 Stud.	Intelligenz: APM von Raven, selektierte Extremgruppen, exp. Bed.: ohne/mit Pfeildiagramm der Variablenbeziehungen	Anzahl Spielperioden mit steigendem Gesamtkapital ("TRENDU")	positiv, tend. sign., keine sign. Interaktion mit exp. Bed. "ohne/mit Pfeildiagramm"
Kreuzig (Intelligenz 1983)	Managementspiel, politische und ökonomische Entscheidungen in einer fiktiven Gemeinde	48 Stud.	nonverbale, kulturfreie Tests: a) CFT3 von Catell b) French et al. c) APM von Raven	Gesamturteil 17 ökonom. u. soziale Größen "Generalgütekriterium"	a) n.s. b) n.s. c) n.s. Kein Interaktionstest

Abbildung 2.17.: Einflüsse der Testintelligenz auf die Effizienz betrieblicher Entscheidungen (Fortsetzung)

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Funke (Dynamische Systeme 1985)*	Schädlingsbekämpfung in einem Biotop	43 Stud.	Experimentalversion des "Berliner Intelligenzbaukastens" von Jäger (Intelligenzstrukturmodell 1982)	Abstand zw. Ziel- und Systemzustand ("Anwendungsgüte")	Keine sign. Zusammenhänge, F-Werte liegen unter Eins
Hussy (Sackgasse 1985)*	Mondlandung: Raumfähre zum Schwebestillstand bringen	60 Stud.	Intelligenz: CFT3 von Weiss; exp. Bed.: 1. Variablenzahl 3 vs. 5, 2. Transparenz: niedrig/hoch variiert über Vollständigkeit des Feedback	Punktescore, berücks. Endgeschwindigkeit und Höhe	n.s. mittlere Korrr.: 0,21 tend. sign. Interaktion mit Variablenzahl, Interaktion mit Transparenz n.s. (Re-Analyse der Korrr. mit z-Transformation)
Hussy/Gran-zow (Determinanten 1985)*	Mondlandung: Raumfähre zum Schwebestillstand bringen	160 Stud.	Gesamtwert des Intelligenztests von Jäger (Intelligenzstrukturmodell 1982); exp. Bed.: ohne/mit Rückmeldung, ohne/mit ext. Gedächtnis, transparent/intransparent (8 Versuchsgruppen)	Punktescore, berücks. Endgeschwindigkeit und Höhe	Testintell. sign. besser, Korrr.: 0,31, unterschiedl. Zusammenhänge bei exp. Bed. aber kein systematisches Muster
Putz-Osterloh (Selbstreflexionen 1985)	Managementspiel, ökologische und wirtschaftl. Entscheidg.	50 Stud.	Verschiedene Maße von Jäger (Intelligenzstrukturmodell 1982): a) numerische Intelligenz b) verbale Intelligenz c) figural-bildhafte Int. d) Verarbeitungskapazität e) Bearbeitungsgeschwdg. f) Einfallsreichtum g) Gesamtwert alle Tests Exp. Bed.: ohne/mit Fragen zur Reflexion des eigenen Vorgehens bei vorangehender Problemlösungsaufgabe	Aufsumm. z-Wert über versch. Krit. (Viehbest., Bevölkerung Hungertote, Weidefläche, Wasser etc., (Maßgröße "Sz")	a) 2 n.s. b) 2 n.s. c) 1 n.s., 1 sign. positiv (0,36) d) 2 n.s. e) 1 n.s., 1 sign. positiv (0,39) f) 2 n.s. g) 1 n.s., 1 sign. positiv (0,36) Kein Interaktions-test (Alle sign. Korrr. entstammen Versuchsgruppe mit Aufddg. zur Reflexion des Vorgehens)

*: Studie wird nur nachrichtlich vermerkt, da kein ökonomisches Effizienzmaß

"Demgegenüber erfordern Intelligenztestitems zu ihrer Bewältigung ein sehr viel engeres Spektrum an Verhaltensweisen. Die Einzelitems stehen in der Regel unverbunden nebeneinander (sind also nicht vernetzt), bleiben als Einzelaufgaben weitgehend überschaubar (sind also nur partiell komplex und intransparent) und verändern sich nicht in der Zeit (verfügen also über keine Eigendynamik). Die Items stehen einzeln sukzessiv und nicht – wie unsere Probleme – simultan zur Bearbeitung an. Es gibt eindeutig richtige oder falsche Lösungen, wobei die Versuchsperson einen sehr hohen Grad an subjektiver Gewißheit über die Richtigkeit ihrer Lösung erlangen kann.

Durch die Vorgabe von Einzelitems entfällt das Problem des Überangebotes an Informationen, das sonst Ordnen, Strukturieren und Reduzieren der Daten notwendigerweise erfordert, um diese im Gedächtnis verfügbar zu halten.

Es müssen weder die kritischen Problemfelder aktiv gesucht (Schwerpunkt-bildung) werden, noch selbstbestimmte Ziele gesetzt werden. Maßnahmen sind in wesentlich engerem Bereich zu suchen; die richtige Dosierung einer Maßnahme entfällt." ⁵⁷⁾

Trotz dieser gewichtigen Unterschiede belegen mehrere Experimente, daß test-intelligentere Versuchspersonen unter bestimmten Bedingungen erfolgreicher sind:

- * Bei **Transparenz** über die Variablenbeziehungen eines simulierten Kleinbetriebes schneiden test-intelligentere Personen signifikant besser ab. ⁵⁸⁾ In der Transparenzbedingung, die der des Intelligenztestes eher entspricht, zeigen sich überdies signifikante Korrelationen mit Meßwerten des Informationsverhaltens, das seinerseits eine positive Beziehung zur Effizienz aufweist. ⁵⁹⁾

- * Falls test-intelligentere Personen zur **Selbstreflexion** ihres Verhaltens trainiert werden, zeigen sie ebenfalls signifikant bessere Leistungen. ⁶⁰⁾ Die mit dem Intelligenztest gemessenen Fähigkeiten zur besseren Informationsverarbeitung sind nach diesem Befund nur ein notwendiges Element, das mit anderen eine Verbindung eingehen muß, um zu einer wirksamen Substanz zu werden.

* Falls das Problem nicht in einen semantischen Kontext eingebunden sondern **abstrakt** formuliert wird, zeigen sich test-intelligentere Personen signifikant überlegen.⁶¹⁾ Test-Intelligenz kann offenbar durch problem-spezifisches Vor-Wissen substituiert werden.

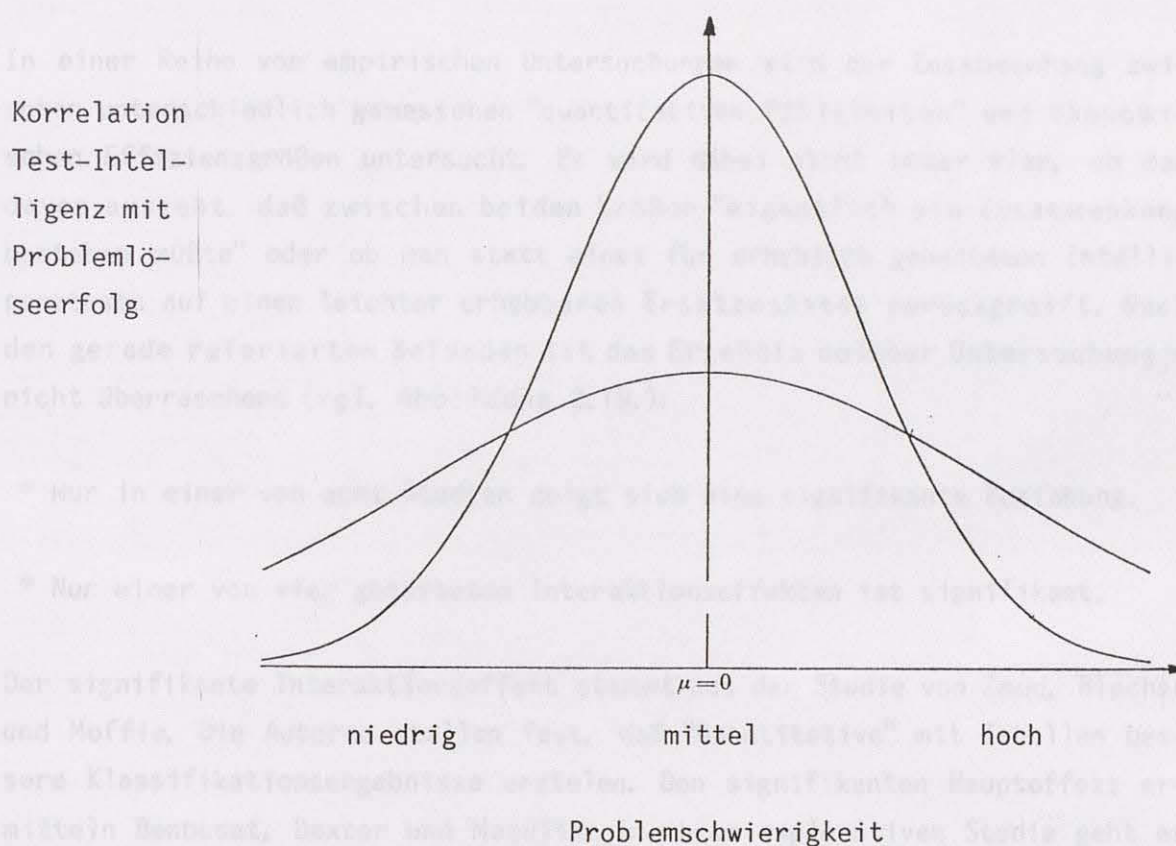
* Falls die Variablenzahl von fünf auf drei gesenkt wird, zeigen testintelligentere Personen signifikant bessere Leistungen.⁶²⁾ Dies kann darauf zurückgeführt werden, daß die überschaubarere Versuchsbedingung mit weniger Variablen stärker den Anforderungen entspricht, die in Intelligenztests gestellt werden.

Die Befunde zeigen eine erstaunliche Parallele zu den Ergebnissen bei der Feldunabhängigkeit, die ja ebenfalls eine bestimmte kognitive **Leistungsfähigkeit** und keine präferierte, selbsteingeschätzte Verhaltensintention mißt. Auch dort stellten wir fest, daß signifikante Leistungsunterschiede nur bei vergleichsweise einfachen Aufgaben erwartet werden können, in denen ähnliche Anforderungen wie beim Meßinstrument gestellt werden.

Hussy faßt die Befunde zur Test-Intelligenz zu der Hypothese zusammen, "daß immer dann, wenn die Problemschwierigkeit sehr niedrig oder sehr hoch ist, die Testintelligenz kein guter Prädiktor für die jeweilige Problemlöseleistung darstellt..."⁶³⁾, während bei einem mittleren Bereich der Problemschwierigkeit "... die Vorhersage der Problemlöseleistung aus der Testintelligenz das Zufallsniveau doch deutlich übersteigen sollte."⁶⁴⁾ Er fügt als "spekulative Aussage" hinzu, "daß es weitgehend gleichgültig ist, durch welche Determinanten der Problemschwierigkeit die Variation erfolgt: es wird immer ein Grad an Schwierigkeit geben, bei dem die Testintelligenz keine Aussage über die Lösungsgüte zuläßt."⁶⁵⁾ (Vgl. hierzu Abbildung 2.18.).

In einer Nachfolgestudie mit einem vergleichsweise einfachen System konnte Hussy diese These nicht erhärten.⁶⁶⁾ Wir halten sie dennoch für eine wertvolle heuristische Anregung, vor allem, um nach den **Ursachen** für eine Reduktion des Zusammenhanges zu forschen.

Abbildung 2.18. Zusammenhang von Test-Intelligenz und Problemlöseerfolg in Abhängigkeit von der Problemschwierigkeit (nach Hussy, Sackgasse 1985)



Fazit: Die in gängigen Tests gemessene Intelligenz ist generell gesehen kein guter Prädiktor für ökonomische Effizienzmaße bei komplexen, dynamischen, partiell intransparenten Problemen. Unter situativen Randbedingungen, die denen der Test-Situation eher entsprechen, oder wenn man die mit Intelligenztests nicht gemessenen Fähigkeiten des Informationsverhaltens wirksam trainiert, können sich signifikante Zusammenhänge ergeben. Wir halten es deshalb beim augenblicklichen Stand der Forschung nicht für angemessen, die mit Intelligenztests erfaßten Eigenschaften als unbeachtlich anzusehen. Wichtiger als diese generelle Einschätzung ist jedoch die gerade erst begonnene Erforschung der **Ursachen** für den unerwartet seltenen Zusammenhang. Diese Ursachenforschung kann unseres Erachtens nur dann Erfolge zeigen, wenn man pfadanalytisch bestimmt, aufgrund welcher Eigenschaften des Informationsverhaltens test-intelligentere Personen höhere Leistungen erzielen. Ansatzpunkte liefern die im Abschnitt 2.4. referierten Arbeiten, namentlich die von Putz-Osterloh und Lür durchgeführten Studien.

2.3.4.2. Quantitative Fähigkeiten: nur selten beachtliche Einflüsse auf die Effizienz

In einer Reihe von empirischen Untersuchungen wird der Zusammenhang zwischen unterschiedlich gemessenen "quantitativen Fähigkeiten" und ökonomischen Effizienzgrößen untersucht. Es wird dabei nicht immer klar, ob man davon ausgeht, daß zwischen beiden Größen "eigentlich ein Zusammenhang bestehen müßte" oder ob man statt eines für erheblich gehaltenen Intelligenztests auf einen leichter erhebbaren Ersatzmaßstab zurückgreift. Nach den gerade referierten Befunden ist das Ergebnis solcher Untersuchungen nicht überraschend (vgl. Abbildung 2.19.):

* Nur in einer von acht Studien zeigt sich eine signifikante Beziehung.

* Nur einer von vier getesteten Interaktionseffekten ist signifikant.

Der signifikante Interaktionseffekt stammt aus der Studie von Zmud, Blocher und Moffie. Die Autoren stellen fest, daß "Quantitative" mit Tabellen bessere Klassifikationsergebnisse erzielen. Den signifikanten Haupteffekt ermitteln Benbasat, Dexter und Masulis. In ihrer explorativen Studie geht es ausschließlich um die Optimierung einer nicht-linearen bivariaten Funktion. Daher ist der Zusammenhang mit mathematischen Fähigkeiten nicht sonderlich überraschend.

Fazit: Die Befundauswertung bestätigt und erhärtet unsere Aussagen zur Test-Intelligenz.

2.3.5. Konsequenzen: Plädoyer für eine differenzierte Forschung

Welche Konsequenzen sind aus den vorliegenden Ergebnissen zu ziehen?

1. Zunächst einmal gilt es anzuerkennen, daß man bei komplexen betriebswirtschaftlichen Entscheidungen nicht von einer generellen Überlegenheit von Entscheidungsträgern mit bestimmten kognitiven Merkmalen ausgehen kann. Insbesondere kann man nicht davon ausgehen, daß test-intelligente

Abbildung 2.19.: Einflüsse quantitativer Fähigkeiten auf die Effizienz betrieblicher Entscheidungen

Quelle	Aufgabe	Subjekt	Persönlichkeitsvariable	Effizienzgröße	Befund
Kozar (Decision Making 1972)	Unternehmensspiel, Beschaffung und Produktion	22 Stud.	Admission Test for Graduate Study in Business, quantitative Fähigkeiten; exp. Bed.: Bildschirm vs. EDV-Liste	Kosten ("Total Cost")	n.s., keine sign. Interaktion mit Präsentationsmedium
Rowland/Gardner (Business Gaming 1973)	Unternehmensspiel, Marketing	200 Stud. in 28 Gruppen	Lösung von vier mathematischen Problemen: a) Anzahl korrekt gelöster Probleme b) Zeitverbrauch	Endwert Eigenkapital ("ending owner's investment")	a) n.s., Korr.: 0,23 b) n.s., Korr.: -0,29
Smith (Inquiry Techniques 1975)	Unternehmensspiel, Beschaffung und Produktion	17 Stud.	Selbsteinschätzung der Vpn bezügl. ihrer Beherrschung quantitativer Techniken	1. Kosten 2. Prognosegenauigkeit	1. n.s. 2. n.s. Kein Interaktionstest
Benbasat/Dexter/Masulis (Human Computer Interface 1981)	Finden einer optimalen Preis-Mengen-Kombination	50 Stud.	Erfahrung mit Computern, korreliert bei Versuchspopulation lt. Autoren mit Ausmaß mathematischer Vorbildung; exp. Bed.: a) Art der Dialogführung, b) Kommandos ohne vs. mit Abkürzungsmöglichkeit	a) Finden der optimalen Kombination b) Erzielter Gewinn	a) Erfahrene sign. häufiger b) Erfahrene sign. besser Kein Interaktionstest
McIntyre (Marketing Models 1982)	simul. Verkaufseinsatzplanung, Aufteilung des Budgets auf verkaufsgebiete	69 Stud.	mathemat. Fähigkeiten, "GMAT"-Punktwert; exp. Bed.: Entscheidungsmodell niedrig/hohe Zufallsschwkg., drei/sechs Verkaufsgebiete	Erzielte Gewinne, relativ zum Optimum	Haupteffekt n.s., keine sign. Interaktion mit Modelleinsatz, übrige Int. nicht getestet
Zmud/Blocher/Moffie (Color Graphic 1983)	Bewertung von Risiken bei Rechnungsprüfungen	51 interne Revisoren	16 mathematische Probleme, Test ähnlich "GMAT"; exp. Bed.: a) Tabelle vs. farbige Balkendiagramme, b) 5 vs. 9 Informationen	Ob Risikoklasse der Rechnung korrekt klassifiziert wird	n.s., durchschn. Korr.: 0,04; (Re-Analyse "Table 5" und "Table 6"), keine sign. Interaktion mit Angebotsmenge (Re-Analyse Tab. 5 u. 6 Vgl. d. Korr.), sign. Interaktion mit Präsentationsform: Quantitative mit Tabelle besser
Henneman/Rouse (Fault Diagnosis Tasks 1984), 1. Exp.	Diagnose von Fehlern in fiktivem, "kontextfreiem" techn. System in Netzwerkdarstellung	34 Stud.	"Mathematical Usage Test" des "American College Testing Service", wird bei Entscheidung über die Aufnahme von Studienbewerbern mit zugrundegelegt	Kosten ersetzter Komponenten ("COST")	n.s., Korr.: 0,17
Henneman/Rouse (Fault Diagnosis Tasks 1984), 2. Exp.	Diagnose von Fehlern in kontextgebundenem System, "aircraft and automobile power plants"	26 Stud.	"Mathematical Usage Test" des "American College Testing Service", wird bei der Entscheidung über die Aufnahme von Studienbewerbern mit zugrundegelegt	Kosten ersetzter Komponenten, ("COST")	n.s., Korr.: 0,02

Personen "un"-intelligenten überlegen seien.

2. Komparative Vorteile bestimmter Typen bei bestimmten Aufgaben sind nur dann zu erwarten, wenn die gemessene Leistungsfähigkeit den Anforderungen der Aufgabe entspricht. Die praktische Relevanz der bisher vorgelegten Ergebnisse ist begrenzt, denn komparative Vorteile Test-Intelligenter, Feldunabhängiger und Reflexiver konnten nur bei vergleichsweise einfachen Aufgaben nachgewiesen werden. Bei den wichtigen komplexen Aufgaben diskriminieren diese Eigenschaften nicht.

3. Besonders kritisch zu beurteilen sind Konstrukte, die lediglich unspezifische, selbsteingeschätzte Verhaltensintentionen erfassen. Aus ihnen folgen weder generelle noch differentielle Leistungsvorteile. Daß man gerade aus solchen Stilen besonders häufig und besonders nachdrücklich Empfehlungen für die Selektion, Aufgabenzuweisung und Schulung von Führungskräften oder die "stilgerechte" Gestaltung von Management-Informations- und Entscheidungsunterstützungssystemen ableitet, und zwar **ohne** die bereits vorliegende empirische Evidenz kritisch und systematisch zu würdigen, belegt nur die Notwendigkeit unserer Bestandsaufnahme.⁶⁷⁾

Man sollte aber nicht in den umgekehrten Fehler verfallen und das Kind mit dem Bad ausschütten, indem man wie z. B. Huber behauptet, daß der "Faktor Mensch" bei der Gestaltung moderner Informationssysteme völlig unerheblich sei, weil diese in Zukunft so flexibel würden, daß sie sich an jegliche Persönlichkeit anpassen werden.⁶⁸⁾ Gerade weil man sich mit dem menschlichen Informationsverhalten befaßt hat, wurde der **Nutzen**, der den höheren Kosten flexiblerer Lösungen gegenübersteht, erkannt.⁶⁹⁾ Ohne die programmatische Betonung Kognitiver Stile wäre die Entwicklung menschengerechter Informationssysteme sicherlich nicht in dem Maße erfolgt. Der technologiezentrierte Monokausalismus ist unseres Erachtens ein genauso gefährlicher Irrweg wie der persönlichkeitszentrierte Monokausalismus. Es gilt vielmehr, das komplexe Zusammenspiel von Aufgabenanforderungen, technologischen Möglichkeiten und Verhaltensantrieben und -barrieren in einer Systemlösung integrativ aufeinander abzustimmen. Nicht das grundlegende Konzept, daß das menschliche Problemlöseverhalten die Qualität des Ergebnisses beeinflusst, ist falsch, sondern der **Weg**, den man bisher beschritten hat.

Dieser Weg sah meistens so aus, daß man eine leicht erhebbare Stilvariable zu einem recht grob gemessenen Informationsverhalten oder einer hypothetisch geäußerten Verhaltensintention in Beziehung setzte, nicht etwa zur Effizienz, wie in den von uns referierten Studien. Aus mäßig ausgeprägten, häufig divergierenden Befunden, wurde die sehr allgemein gehaltene Forderung abgeleitet, daß man Persönlichkeitsunterschiede bei der Gestaltung von subjektgerechten Informationssystemen berücksichtigen sollte. Hinter dieser Forderung steht die Überlegung, daß verschiedene Wege zum gleichen Ziel führen, daß aber bestimmte Menschen nur bestimmte Wege beschreiten wollen oder können und man ihnen diese Wege nicht abschneiden dürfe.

Auch wenn man dieser Überlegung folgt, muß man den Nutzen der durchgeführten Studien sehr in Frage stellen: Es fehlt erstens der Nachweis, daß **bestimmte** Persönlichkeitsmerkmale das tatsächliche Informationsverhalten zufriedenstellend erklären. Es fehlt zweitens der Beleg, daß die untersuchten Verhaltensmeßwerte oder -intentionen tatsächlich für die Effizienz beachtlich sind. Was aber vor allem fehlt, ist eine **pfadanalytische Prüfung** der Wirkungskette: Persönlichkeitsvariable - tatsächliches Problemlöseverhalten - Problemlöseerfolg. Eine solche hypothesengeleitete Prüfung ist unseres Wissens noch nicht vorgenommen worden.⁷⁰⁾ Bezüglich der Untersuchungen aus der amerikanischen Managementforschung kann man nur feststellen, daß die vom Kognitiven Stil abhängigen Variablen entweder zu schwach beeinflußt werden, nur schwach auf die Effizienz einwirken, oder daß ihre effizienzwirksamen Pfade durch andere kompensiert werden. Da das intervenierende Problemlöseverhalten aber gerade nicht berücksichtigt wurde, bleibt die relative Bedeutung der einzelnen Möglichkeiten reine Spekulation. Es nützt sehr wenig, wenn auf der einen Seite festgestellt wird, daß bestimmte Kognitive Stile zu Unterschieden im Informationsverhalten führen, auf der anderen Seite aber bei komplexen betriebswirtschaftlichen Problemen weder signifikante Haupt- noch Interaktionseffekte mit der Effizienz nachgewiesen werden können. Notwendig ist vielmehr eine **simultane** Betrachtung der Merkmalsblöcke Person, Aufgabe, Informationssystem, Informationsverhalten und Effizienz.

Hierzu bedarf es des folgenden Forschungsprogramms:

1. Der Ausgangspunkt der Betrachtung ist die **Aufgabe**, nicht der Persönlichkeitstest, der meist für ganz andere Zwecke geschaffen wurde. Man muß klären, welche **Anforderungen** eine bestimmte Aufgabe stellt. Dies ist ein substantielles betriebswirtschaftliches Problem, das bisher nur bei der Anwendung formaler Entscheidungshilfen unter dem Stichwort "Prämissenkritik" ernsthaft aufgegriffen worden ist. Wegen der Komplexität der Aufgaben dürfte es sich empfehlen, diese in Teilaufgaben zu zerlegen und für jede Teilaufgabe spezifische Anforderungsprofile zu erarbeiten. Trotzdem sollte die Gesamtaufgabe und nicht etwa eine überschaubare Teilaufgabe vorgegeben werden, denn es kommt darauf an, die Effizienzwirkungen der **subjektiven** Abgrenzung und Faktorisierung eines komplexen Problems zu untersuchen.
2. Sodann sind **Effizienzmaßstäbe** zu erarbeiten – und zwar sowohl für die Gesamtaufgabe als auch für die Teilaufgaben – um entscheiden zu können, worauf mögliche Insuffizienzen beruhen. Die bisher verwendeten Effizienzmaße sind unseres Erachtens zu global, um eine **differenzierte Insuffizienzanalyse** zu ermöglichen, die eine Voraussetzung für eine qualifizierte Bewertung von Schulungsmaßnahmen, Kognitiven Stilen oder Bedingungen des Informationssystems darstellt.
3. Bei den Aufgaben und Effizienzmaßstäben ist zu überprüfen, ob es **konsistente interpersonelle Leistungsunterschiede** gibt. So stellen beispielsweise Elstein/Shulman/Sprafka⁷¹⁾ und Pankoff/Virgil⁷²⁾ fest, daß die Leistungsfolgen derselben Personen bei Wiederholungen formal vergleichbarer Aufgaben unterschiedlich ausfallen. Wenn es aber keine **reproduzierbaren** interpersonellen Leistungsunterschiede gibt, dann kann es auch keine (reine) Frage der "Persönlichkeit" sein, ob ein Individuum bei einem bestimmten Problem besser oder schlechter abschneidet als ein anderes. Man muß also erst einmal eine hinreichende Menge gleichartiger Probleme lösen lassen, um eine zuverlässige mittlere Leistungsschätzung zu gewinnen. Auch eine Bestimmung vergleichbarer Aufgaben setzt eine gründliche Aufgabenanalyse voraus. Ob die Aufgaben tatsächlich homogene Anforderungen stellen, ist anschließend empirisch zu prüfen.⁷³⁾

4. Als viertes muß das **intervenierende Informationsverhalten** prozeßbegleitend und differenziert erhoben und zu Meßwerten verdichtet werden, von denen man aufgrund theoretischer Überlegungen annimmt, daß sie eine positive Beziehung zur Effizienz aufweisen. Sodann ist diese Vermutung zu prüfen, gegebenenfalls sind besser geeignete Meßwerte zu entwickeln. Erst wenn man weiß, welche Informationsaktivitäten tatsächlich die Effizienz positiv beeinflussen, kann man versuchen, diese zu fördern; erst wenn man weiß, welche Aktivitäten dysfunktional wirken, kann man versuchen, diese zu verhindern. Dabei reicht es nicht, nur die Informationsbeschaffung prozeßbegleitend zu erfassen und diese schon als eine "Effizienzgröße" anzusehen, wie z. B. in den Experimenten der Minnesota Studies.⁷⁴⁾ Es kommt vielmehr auch auf die **Zusammenhänge** zwischen verschiedenen Aktivitäten an. Und auch beim Informationsverhalten stellt sich die Frage, wie stabil oder variabel es bei unterschiedlichen Aufgaben ausfällt.
5. Erst an dieser Stelle erhebt sich die Frage, **welche Persönlichkeitstests** dieses Verhalten erklären können. Dabei darf man wiederum nicht von leicht erhebbaren Maßen ausgehen. Maßgeblich ist vielmehr, ob der auszuwählende oder zu konstruierende Test die Fähigkeiten mißt, die für die Lösung des untersuchten (Teil-) Problems beachtlich scheinen.
6. Gestützt auf dieses theoretische und empirische Wissen kann man dann daran gehen, die Wirkungen von kognitiven Persönlichkeitsmerkmalen und deren mögliche Interaktionen mit Bedingungen des Informationssystems **pfadanalytisch** zu erforschen.

Dies ist in der Tat ein aufwendiges, langfristig angelegtes, aber notwendiges Forschungsprogramm, wenn es um Erkenntnis und nicht um die Anhäufung fragwürdiger empirischer Befunde gehen soll. Sein Kerngedanke ist die Analyse des menschlichen Informationsverhaltens und seiner Effizienzwirkungen, nicht das Vertrauen in die Leistung bestimmter Persönlichkeitstests. Wir können uns gut vorstellen, daß man ein effektives Informations-System auch ohne Beachtung der bisherigen Forschungsergebnisse über "Kognitive Stile" konstruieren kann. Ob man aber ohne Effizienzhypothesen über das Verhalten der Benutzer auskommen kann, möchten wir bezweifeln.

2.4. INFORMATIONSVERVERHALTEN UND EFFIZIENZ

2.4.1. Überblick

2.4.1.1. Abgrenzung der untersuchten Studien

Entgegen einer häufig geäußerten Vermutung gibt es bereits eine ganze Reihe empirischer Studien, in denen man das Informationsverhalten bei komplexen Problemen prozeßbegleitend erfaßt:

- * In den Arbeiten von Schroder und Mitarbeitern¹⁾ sowie Witte und seinen akademischen Schülern wird ein **schriftlich** niedergelegtes Informationsverhalten systematisch **inhaltsanalytisch** ausgewertet.²⁾ Im Mittelpunkt von Wittes vielbeachteten Forschungsarbeiten stehen Erscheinungsformen, Ursachen und Wirkungen der "unvollkommenen Informationsnachfrage".³⁾
- * Zahlreiche Untersuchungen über die Nutzung von Management-Informations- und Entscheidungs-Unterstützungs-Systemen verfolgen das Informationsverhalten mit Hilfe von **Logbüchern**, die vom Computer verwaltet werden.⁴⁾
- * In der psychologischen Entscheidungsforschung, im Marketing und in der verhaltensorientierten Rechnungswesenforschung gibt es eine exponentiell wachsende Zahl von Untersuchungen, in denen verschiedene Varianten von **Informations-Display-Techniken** verwendet werden.⁵⁾ Während man in älteren, teilweise bereits in den fünfziger Jahren durchgeführten Studien Lochkarten zieht oder einfache Apparaturen betätigt, setzt man in neueren Untersuchungen computergestützte Displays ein.⁶⁾
- * In den gleichen Forschungsdisziplinen werden Informationsprozesse auch mit Hilfe der **"Denke-Laut-Technik"** erforscht.⁷⁾
- * In neueren Arbeiten der Problemlöseforschung werden nicht nur einfache Informationsprobleme, wie z. B. das zufällige Betrachten einer Werbeanzeige,⁸⁾ sondern auch das Lösen komplexer geistiger Aufgaben mit **Blickaufzeichnungen** untersucht.⁹⁾

* Schließlich gibt es Studien, die verbalisiertes Informationsverhalten mittels **sprachstatistischer Techniken** auswerten.¹⁰⁾ Erste Ergebnisse dieser neuen Richtung weisen auf systematische Unterschiede zwischen erfolgreichen und erfolglosen Problemlösern hin.¹¹⁾

Doch nicht nur die Datenerhebung, auch die **Auswertung von Prozeßdaten** hat erhebliche Fortschritte gemacht. Es gibt bereits eine Reihe von konkurrierenden und komplementären Ansätzen, um Verlaufsdaten zu quantitativen Meßwerten zu verdichten.¹²⁾

Die Bestandsaufnahme der Studien zum Informationsverhalten fällt besonders schwer. Bei den bisher referierten Arbeiten über die Wirkung von Informationssystemen und Persönlichkeitsmerkmalen ist trotz aller Unterschiede bei der Operationalisierung der Variablen ein gemeinsames Leitkonzept zu erkennen, dem sich die Autoren verpflichtet fühlen. Demgegenüber haben wir es jetzt mit Beiträgen aus mehreren wissenschaftlichen (Teil-) Disziplinen zu tun, die praktisch keinen Bezug aufeinander nehmen. Das allen gemeinsame Konstrukt "Informationsverhalten" wird verschiedenartig definiert und gegliedert. Die gerade erwähnten Meßverfahren sind bei näherem Hinsehen keine Alternativen. Sie erfassen vielmehr unterschiedliche Dimensionen des Informationsverhaltens. Hinzu kommt, daß man auch bei diesen Studien heterogene Aufgaben, Effizienzmaßstäbe und Erhebungs- und Auswertungstechniken verwendet. Das Problem der Vergleichbarkeit wird nur dadurch gemildert, daß wir uns auf die vergleichsweise kleine Teilmenge dieser Studien beschränken, die das prozeßbegleitend gemessene Informationsverhalten zur **Effizienz** in Beziehung setzen.

Diese Teilmenge kann man in drei Gruppen von Studien untergliedern, die ähnliche Aufgaben betrachten und mit vergleichbaren Methoden analysieren:

- a) Feldstudien von Entscheidungen der Geschäftsführung,
- b) experimentelle Studien mit computergestützten Managementspielen,
- c) experimentelle Analysen von Beurteilungs- und Prognoseproblemen,

Zu a) Feldstudien

Bei der ersten Gruppe spielen **soziale Interaktionsprozesse** eine ausschlaggebende Rolle. Das Informationsverhalten beschränkt sich nicht auf den Abruf von Informationen von einem Display, einem maschinellen Informationssystem oder einem neutralen Versuchsleiter. Es handeln vielmehr interessengeleitete Informanden, die ihre Adressaten durch selektive Lieferung, Bewertung, geschickte Präsentation und Filterung von Informationen aktiv beeinflussen wollen. Die Informanden werden aber auch von ihren Adressaten beeinflusst. Zur Erklärung der Effizienz dieser Prozesse ist eine **Interaktionsbetrachtung** erforderlich.

Die Arbeiten dieser Gruppe unterscheiden sich auch hinsichtlich der untersuchten Subjekte und der verwendeten Methoden zur Messung des Informationsverhaltens von den übrigen Studien: Mehrheitlich werden Top-Manager untersucht, die strategische Entscheidungen bearbeiten. Das Informationsverhalten kann meist nur **partiell** mittels systematischer **Inhaltsanalysen** von Protokollen, Gesprächsnotizen oder Korrespondenzunterlagen, sowie schriftlichen und mündlichen **Befragungen** des nachträglich erinnerten Verhaltens gemessen werden.

Zu b) Computergestützte Managementspiele

Computergestützte Simulationen von Managementproblemen werden sowohl von Betriebswirten als auch von Psychologen benutzt, um die Effizienzwirkungen des Informationsverhaltens bei komplexen, eigen-dynamischen Problemen zu untersuchen. Diese Simulationen unterscheiden sich von den nachfolgenden Beurteilungs- und Prognoseproblemen dahingehend, daß die "Umwelt" in Form computermodellierter, stochastischer Reaktionsfunktionen oder menschlicher Wettbewerber auf eigene Aktivitäten reagiert und man daraufhin **erneut**, auf Basis des nun geschaffenen Zustandes, agieren muß. Sie stellen die Anforderung zu **lernen**, wie das System oder die Wettbewerber auf eigene Maßnahmen reagieren. Hierzu muß man nicht nur Aktionsparameter festlegen, sondern auch **Ergebnisse** kritisch **analysieren** und angemessene **Hypothesen** über Art, Wechselwirkungen, Stärke und Verlauf von funktionalen Beziehungen **bilden**, um wohldosierte Maßnahmen treffen zu können.

Da sich die meisten Studien auf **individuelle Problemlöseprozesse** konzentrieren, fehlt die direkte soziale Interaktion mit dem Mitspieler: Es gibt weder Möglichkeiten zur Verhandlung noch zum gemeinsamen Entwurf neuer Lösungsalternativen. Die Aktionsparameter sind vielmehr durch die Spielregeln gegeben. Ein weiterer Unterschied zur ersten und dritten Gruppe besteht darin, daß man meistens **studentische Versuchspersonen** analysiert. Zur Messung des Informationsverhaltens wird überwiegend die Technik der "**Denke-Laut-Protokolle**" eingesetzt.

Zu c) Beurteilungs- und Prognoseprobleme

Die ausgewerteten Studien über Beurteilungs- und Prognoseprobleme unterscheiden sich von den bisherigen dadurch, daß nur eine **einstufige Entscheidungssituation** betrachtet wird. Der Entscheidungsträger holt Informationen ein, verarbeitet diese unter Berücksichtigung seines bisherigen Wissens und gibt daraufhin ein Urteil oder eine Prognose ab, oder er trifft eine einmalige Wahl, gestützt auf seine Prognose der mutmaßlich besten Alternative. Er muß aber nicht mit den Konsequenzen seiner Entscheidungen weiterleben und auf eine Umweltreaktion seinerseits reagieren. Es handelt sich insofern um ein **statisches Entscheidungsproblem**.

Ein weiteres Unterscheidungsmerkmal betrifft die **Art der zu ergreifenden Maßnahmen**: Während bei den Simulationsstudien zahlreiche Maßnahmen in einer Strategie zu kombinieren sind und bei den Feldstudien die Alternativen (mit) entwickelt oder verhandelt werden, wird bei den Beurteilungs- und Prognoseproblemen nur eine **einwertige Ergebnisabschätzung** verlangt.¹³⁾ Mit dieser Festlegung **entfallen** auch die in den Feldstudien strategischer Entscheidungen auftretenden **Probleme der Zielbildung**: Es gibt keine Unklarheiten über Zielobjekt, -eigenschaften und -maßstäbe und keine Konflikte hinsichtlich der Gewichtung.

Wir vermuten, daß bei dieser Gruppe von Studien die **Informationsaktivitäten eine geringere Rolle spielen** als bei den computergestützten Managementspielen: Erstens besteht ein geringerer Informationsbedarf, weil weniger Aktionsparameter festzulegen sind und nur eine einmalige Beurteilung, Pro-

gnose oder Wahl zu treffen ist. Zweitens können die bei diesen Studien untersuchten Praktiker die Informationssuche durch vorhandenes **Expertenwissen** substituieren.

2.4.1.2. Die Hypothesen

Eine Bestandsaufnahme der empirischen Ergebnisse zu den Effizienzwirkungen des Informationsverhaltens setzt voraus, daß man die Befunde bestimmten Hypothesen gegenüberstellt. In diesem Punkt ist ein erhebliches Defizit der bisherigen Forschung festzustellen. Die meisten Untersuchungen gehen rein explorativ vor. Sie bleiben weit hinter den theoretischen Differenzierungen zurück, die Witte bereits 1972 vorgenommen hat.¹⁴⁾ Deshalb beschränken wir uns auf einige Leithypothesen, denen die vorgenommenen Operationalisierungen mehr oder weniger angemessen zugeordnet werden können. Dies sind:

- a) die Baustein-Hypothese,
- b) die Erkenntnis-Hypothese,
- c) die Steuerungs-Hypothese und
- d) die Verarbeitungs-Hypothese

Zu a) Baustein-Hypothese

Die Baustein-Hypothese betrachtet Informationen als Rohstoffe, die zur Produktion von Entscheidungen benötigt werden. Sie geht davon aus, daß der Entscheidungsträger diese Rohstoffe noch nicht besitzt, sondern sich erst extern beschaffen muß. In der einfachsten Version behauptet sie:

Je mehr Daten beschafft werden, desto höher die Effizienz der Entscheidung.

Eine differenziertere Version verlangt, daß die nachgefragten Bausteine zur Problemlösung besonders geeignet und für den Entscheidungsträger neuartig sind. Hier lautet die Hypothese:

Je mehr "relevante"¹⁵⁾, d. h. problembezogene und subjektiv neuartige Daten beschafft werden, desto höher die Effizienz der Entscheidung.

Die Effizienzvermutung gilt bei dieser Variante nur bis zu dem Punkt, an dem der Informationsgrad Eins beträgt, d. h. die vorhandene gleich der notwendigen Information ist.¹⁶⁾

Eine dritte Version greift auf die gleiche theoretische Basis wie die Experimente zur sog. Informations-"Über"-Lastung zurück, die wir im Teil 2.2. referiert haben. Sie geht von sinkendem Grenznutzen und steigenden oder konstanten Grenzkosten der Informationsbeschaffung aus und postuliert:

Mit zunehmender Menge an beschafften Informationen nimmt die Effizienz der Entscheidung zunächst zu und dann ab.

Zur Prüfung der anspruchslosen Variante der Baustein-Hypothese können alle Studien herangezogen werden, in denen die Menge an "nachgefragten", "beschafften", "gesuchten" oder "aufgenommenen" Informationen zur Effizienz in Beziehung gesetzt wird. Die Prüfung der zweiten Variante setzt eine Qualifizierung der beschafften Bausteine nach ihrer Problem- und Subjektrelevanz voraus. Zur Beurteilung der dritten Hypothesenversion müssen nicht-lineare Funktionen geschätzt werden. Da die meisten Untersuchungen nur den Ansprüchen der ersten Variante genügen, muß sich unsere Bestandsaufnahme weitgehend auf die Bewertung dieser recht pauschalen Aussage beschränken.

Zu b) Erkenntnis-Hypothese

Die Artikulation einer Informationsnachfrage setzt einen geistigen Prozeß voraus, in dessen Verlauf ein Problem strukturiert und konturiert wird.

Hierzu schreibt Witte:

"Wenn jemand in der Lage ist, klug zu fragen, dann weist er dadurch nach, daß er bereits Einsicht in die Problemstruktur besitzt. Er hat eine gewisse intellektuelle Reife gewonnen, indem er ein vereinfachtes Abbild der Entscheidungsaufgabe erarbeitet und erste Lösungsansätze gefunden hat. Diese bereits vollzogene Rohanalyse befähigt ihn, eine intensivere Nachfrage-Aktivität zu entfalten, als eine Entscheidungs-Person, die den Erkenntnisvorgang nicht durchlaufen hat."¹⁷⁾

Zur Prüfung dieser Hypothese könnten alle Studien herangezogen werden, in denen man die Informationsnachfrage, d. h. die Artikulation von Informations-**Wünschen** zur Effizienz in Beziehung setzt. Es ist nicht notwendig, die tatsächlich beschafften Informationen zu betrachten, denn die Erkenntnis-Erklärung behauptet auch dann einen positiven Effekt, wenn **keine** Informationslieferung erfolgt.

Auf der anderen Seite hebt Witte mit der Umschreibung "klug" eindeutig auf den **Qualitätsaspekt** der Nachfrage ab. Es wäre daher unangemessen, der Hypothese Befunde gegenüberzustellen, in denen lediglich Mengen von Nachfrageaktivitäten ausgezählt werden. Hypothesengerechte Prüfungen fordern vielmehr, daß man die Qualität des Problemwissens der Entscheidungsträger zur Effizienz in Beziehung setzt. Wir formulieren daher folgende Hypothese:

Je mehr Informationsaktivitäten ergriffen werden, die eine Erkenntnis des Entscheidungsträgers anzeigen, desto höher die Effizienz der Entscheidung.

Zum Test dieser Behauptung kann auf mehrere Studien zurückgegriffen werden, in denen bestimmt wird, ob der Entscheidungsträger **angemessene Hypothesen** über den Zusammenhang von Variablen artikuliert, denn man geht mittlerweile davon aus, daß sich **Experten** vor allem dadurch auszeichnen, daß sie über ein besonders differenziertes Beziehungswissen über einen bestimmten Gegenstandsbereich verfügen.¹⁸⁾

Zu c) Steuerungs-Hypothese

Die Steuerungs-Erklärung setzt an der Lenkung des Informationsversorgers an. Mit Blick auf die Mensch-Maschine-Kommunikation führt Möllhoff an:

"Die Artikulation von Nachfrage bewirkt eine **bedarfsgerechte Versorgung**: Der Entscheidungsträger erhält ausschließlich jene Informationen, die seinen individuellen und situativen Informationsbedarf befriedigen. Die Artikulation von Nachfrage bewirkt zudem eine **kapazitätsgerechte Versorgung**: Sie steuert den Informationstransfer von dem System zum Individuum gemäß der zeitlichen Disposition des Entscheidungsträgers, der individuellen Informationsverarbeitungsfähigkeit und der individuellen Bereitschaft des Entscheidungsträgers, Informationen zu verarbeiten."¹⁹⁾

Möllhoff fügt hinzu, daß die Informationsnachfrage auch festlegt, welche Informationen geliefert werden und in welcher Reihenfolge die Informationslücken geschlossen werden. In dieser Variante kann die Hypothese nicht geprüft werden, weil keine geeigneten Operationalisierungen vorliegen.

In unserer Bestandsaufnahme kann nur eine auf **soziale** Interaktionsprozesse bezogene Steuerungs-Erklärung beurteilt werden. Sie geht davon aus, daß Versorgung und Nachfrage von Informationen aufeinander abgestimmte Prozesse sein sollen:²⁰⁾

Je höher der Grad der Entsprechung von Informationsnachfrage und Informationsversorgung, desto größer die Effizienz der Entscheidung.

Zur Prüfung dieser These können geeignete Operationalisierungen aus den referierten Feldstudien herangezogen werden.

Zu d) Verarbeitung-Hypothese

Der Schluß von der Informationsbeschaffung auf die Effizienz der Entscheidung ist recht kühn: Welcher Betriebswirt würde schon von Menge und Qualität des Inputs auf die Effizienz eines Betriebes schließen? Daher kann wohl nur bei einfachen Problemstellungen und homogenen Informationsverarbei-

tern²¹⁾ von einer engen Beziehung zwischen Informationsbeschaffung und Effizienz ausgegangen werden. Eine Versorgung mit hinreichend vielen, problemrelevanten Daten stellt unseres Erachtens nur eine **notwendige** Bedingung dar, denn **ohne** Informationen kann im allgemeinen keine gute Entscheidung getroffen werden. Auch ein Experte, der über ein großes "Hintergrundwissen" verfügt, benötigt zur Lösung eines konkreten Problems zusätzliches Faktenwissen, um die "Wenn"-Komponenten seines Beziehungswissens auf faktische Gültigkeit zu prüfen.

Aus diesem Grunde ist es zweckmäßig, eine eigenständige Hypothese aufzustellen, die eine positive Wirkung der Informationsverarbeitung postuliert:

2.4.2.1. Methodische Darstellung der Studien

Je höher das Ausmaß und die Qualität der Informationsverarbeitung, desto höher die Effizienz der Entscheidung.

Feldstudien zurückgegriffen werden (vgl. Abbildung 2.4.2).

Welche Operationalisierungen geeignet sind, Qualität und Ausmaß der Informationsverarbeitung zu erfassen, wird anhand der einzelnen Studien erörtert.

* die Studie von Bringer, Hoffmann und Jütt zur strategischen Planung.

* die Arbeit von 2.4.1.3. Zur Methodik der Bestandsaufnahme

In der nachfolgenden Bestandsaufnahme werden wir ausführlicher als bisher auf einzelne Untersuchungen eingehen. Zusammenfassende Aussagen über die Ergebnisse werden wir seltener und weniger streng formalisiert als bisher vornehmen können. Diese Änderung der Berichterstattung wird zum Teil durch die Heterogenität der Studien erzwungen, zum Teil ist sie auf die größere Relevanz einzelner Studien für unsere Arbeit zurückzuführen. Insgesamt gesehen dient die Bestandsaufnahme mehr dem Zweck, Hypothesen-Elemente für die eigene Arbeit zu entdecken und zu belegen als bereits vorhandene Theorien meta-analytisch zu prüfen.

* Hauschildt und Mitarbeiter werden Studioprofokolle von Geschäftsleuten

Einer rigorosen Meta-Analyse stehen nicht nur der unbefriedigende Stand der Theorie und die divergierenden, meist recht groben Operationalisierungen des Informationsverhaltens im Wege. Schon die Bestimmung der "Zähleinheit" macht Probleme: Wenn in einer typischen Studie über die Wirkung des Infor-

mationssysteme wie der von Benbasat²²⁾ mehrere Angebotsbedingungen experimentell in einem faktoriellen Design orthogonal variiert werden, dann kann man jede getestete Beziehung als ein unabhängiges Ergebnis referieren. Wenn jedoch in einer Studie wie der von Putz-Osterloh und Lürer²³⁾ mehrere voneinander abhängige Informationsaktivitäten zur Effizienz in Beziehung gesetzt werden, dann ist es problematisch, die univariat getesteten Zusammenhänge als unabhängige Beobachtungen anzusehen.²⁴⁾

2.4.2. Ergebnisse der Feldstudien

2.4.2.1. Methodische Beurteilung der Studien

Zur Beurteilung der Effizienz des Informationsverhaltens kann auf fünf Feldstudien zurückgegriffen werden (vgl. Abbildung 2.20.):

- * die Untersuchung **Wittes** zur **erstmaligen Beschaffung von EDV-Anlagen**,
- * die Studie von **Grinyer/Norburn** und **Nutt** zur **strategischen Planung**,
- * die Arbeit von **Lucas** über **Marketinginformationssysteme** und
- * die Analyse von **Entscheidungen der Geschäftsführung durch Hauschildt und Mitarbeiter**.

Gemessen an der Komplexität der untersuchten Entscheidungen kann das Informationsverhalten nur partiell und recht undifferenziert erhoben werden:

- * Witte beschränkt sich auf eine "Spiegelbild"-Erhebung der bei den EDV-Herstellern gespeicherten Korrespondenzakten,
- * Hauschildt und Mitarbeiter werten Sitzungsprotokolle von Geschäftsleitungskonferenzen aus, die zumindest bei einigen Entscheidungen nur die letzte Phase des Entscheidungsprozesses erfassen.

Abbildung 2.20.: Einflüsse des Informationsverhaltens auf die Effizienz betrieblicher Entscheidungen (Feldstudien)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Witte (Informationsverhalten 1972)	erstmalige Beschaffung von EDV-Anlagen,	227 Organisationen	a) Anteil Informations-Nachfrage-Aktivitäten an der Menge aller Aktivitäten eines Entscheidungsprozesses b) absolute Differenz von Nachfrage- und Versorgungs-Aktivitäten, beide relativiert an der Menge aller Aktivitäten eines Entscheidungsprozesses	Innovationsgrad, Expertenurteil, das verschiedene Gütekriterien berücksichtigt	a) n.s. (positiver Zusammenhang, sign. Untersch. zw. Extremgruppen) b) positiv sign. (Detailanalyse zeigt, daß positiver Zusammenhang nur für Versorgung größer als Nachfrage gilt)
Grinyer/Norburn (Planning 1975)	Strategische Planungsprozesse	21 Unternehmen	a) Anzahl Informationen, die verwendet werden, um eine Überprüfung der Strategie zu veranlassen ("used to signal the need for a review") b) Anzahl der Informationen, die bei einer Überprüfung der Strategie verwendet werden c) Anteil der verwendeten Informationen, bezogen auf die empfangenen Informationen	Rendite vor Zins und Steuern, im Vergleich zum jeweiligen Markt des Unternehmens, (z-Werte)	a) n.s. (Korr. 0,19) b) pos., tend.sign. (Korr. 0,41) c) pos.sign. (Korr. 0,51 lt. Angabe im Text S. 82)
Lucas (Information Systems 1975) Sales Force Study, Befund S. 90, vgl. auch Lucas (Performance 1975)	Feldstudie von Verkaufsmanagern, die kleinere und mittlere Kunden betreuen ("sales representatives")	104, 43 und 24 Manager je Division	Intensität der Nutzung verschiedener Berichte: a) im Gespräch mit Kunden b) für Detailanalysen der Geschäftsbeziehung c) Soll-Ist-Abweichungen ("overall progress") d) Verändg. zum Vorjahr ("summary reports") e) Planung von Verkaufsaktivitäten ("planning") f) Stornierungen g) zusätzlich eigene Aufzeichnungen ("records") h) Berechnungen mit Daten (Fragebögen, Wortlaut der Fragen nicht veröffentl.)	Brutto-Auftrags-Eingänge, ("gross bookings")	Schrittweise Regressionen, getrennt für drei Divisionen, bivariate Befunde nicht berichtet a) alle drei n.s. b) alle drei n.s. c) alle drei n.s. d) alle drei n.s. e) alle drei n.s. f) alle drei n.s. g) 1 sign. positiv, (N=104 Fälle), n.s. h) alle drei n.s. Wegen ungünstigem Verhältnis Fallzahl zu Variablen (insg. 17 bzw. 18) sind Ergebnisse zurückhaltend zu bewerten

Abbildung 2.20.: Einflüsse des Informationsverhaltens auf die Effizienz betrieblicher Entscheidungen (Feldstudien)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Lucas (Information Systems 1975) Sales Force Study, Befund S. 101, vgl. auch Lucas (Performance 1975)	Feldstudie von Verkaufsmanagern, die Großkunden betreuen ("account executives")	41 bzw. 22 Manager je Division	Intensität der Nutzung verschiedener Berichte: a) im Gespräch mit Kunden b) für Detailanalysen der Geschäftsbeziehung c) Soll-Ist-Abweichungen ("overall progress") d) Verändg. zum Vorjahr ("summary reports") e) Planung von Verkaufsaktivitäten ("planning") f) Stornierungen g) zusätzlich eigene Aufzeichnungen ("records") h) Berechnungen mit Daten (Fragebögen, Wortlaut der Fragen nicht veröffentlicht)	Brutto-Auftrags-eingänge, ("gross bookings")	Schrittweise Regressionen, getrennt für zwei Divisionen, bivariate Befunde nicht berichtet a) beide n.s. b) beide n.s. c) beide n.s. d) beide negativ, davon 1 sign. e) beide n.s. f) 1 sign. pos., 1 n.s. g) beide n.s. Wegen ungünstigem Verhältnis Fallzahl zu Variablen (insg. 17 bzw. 18) sind Ergebnisse zurückhaltend zu bewerten
Hauschildt (Effizienz 1983), vgl. auch Gemünden/Hauschildt (Top Management Decisions 1985)	Entscheidungen der Geschäftsführung	6 Top-Manager 83 Entscheidungen	a) Anzahl nachgefragter Informationen während der Sitzung des Führungsgremiums b) Anzahl diskutierter Alternativen während der Sitzung des Führungsgremiums	Entscheidungsqualität im ex-post-Urteil der beteiligten Top-Manager	a) n.s. (positiv, eliminiert man die Fälle, die wegen Zielwandels ineffizient sind, ist die Beziehung tend. sign. b) positiv sign. (bei 2 vs. 3 Alternativen, 3 vs. 4 negativ, aber nicht mehr testbar)
Nutt (Process Archetypes 1984)	Strategische Planungsprozesse	63 Organisationen	Fünf versch. Prozeßtypen, die sich durch die ergriffenen Aktivitäten a. Problemdefinition, b. Konzeptualisierung, c. Detailierung, d. Bewertung, e. Implementierung unterscheiden: 1. Historisches Modell: a-c-e 2. Ausgehen von externem Vorschlag: a-c-d-e 3. Beurteilung eines fertigen Plans: a-d-e 4. Ungerichtete Suche: a-e 5. Novative Lösung: a-b-c-d-e	Entscheidungsqualität im ex-post-Urteil der beteiligten Top-Manager	tend. sign. Unterschiede zw. Prozeßtypen, erkl. Var.: 3,3 v. H. (Re-Analyse F-Wert S.228), sign. Unterschiede Typen 1 vs. 2, 3 vs. 4, 5; Detailanalyse zeigt bei den erfolgreichen Typen besonders intensive Informations-Verarbeitung, unterstützt durch formale Planungstechniken

- * Grinyer/Norburn und Nutt führen strukturierte Interviews durch, erfassen also ein nachträglich erinnertes Informationsverhalten,

- * Lucas stützt sich auf schriftliche Fragebögen, die noch größere Validitätsprobleme aufwerfen als persönliche Interviews.

Andererseits:

- * erweisen sich die bei den EDV-Herstellern gespeicherten Akten als reichhaltiger und homogener als die bei den Käufern archivierten Schriftstücke. Berücksichtigt werden nur Entscheidungsprozesse mit vollständigen und "voll brauchbaren" Akten. Die Dokumente werden mittels einer eigens entwickelten Variante der Inhaltsanalyse von zwei Erhebern systematisch und differenziert erhoben.

- * werden bei Grinyer und Norburn durchschnittlich 4,5 Auskunftspersonen pro Unternehmen befragt. Nur diejenigen Tatbestände, bei denen eine auf dem 5%-Niveau statistisch signifikante Übereinstimmung herrscht, werden der Auswertung zugrundegelegt.

- * werden auch bei Nutt mehrere Personen zum gleichen Tatbestand interviewt. Die daraufhin entwickelten Prozeßrekonstruktionen werden mit den Befragten sorgfältig abgestimmt und bereinigt.²⁵⁾

- * informieren die von Hauschildt und Mitarbeitern ausgewerteten Protokolle detailliert über die während der Sitzung erfragten Informationen, diskutierten Alternativen und getroffenen Entschlüsse. Ihre Zuverlässigkeit wird dadurch unterstrichen, daß die Protokolle unmittelbar nach Fertigstellung den Besprechungsteilnehmern zur Bestätigung bzw. Vornahme von Korrekturen zugestellt wurden.

- * Lediglich in der Studie von Lucas gibt es keine Hinweise auf Reliabilität und Validität der nur sehr allgemein beschriebenen Fragen.

In allen fünf Studien wird außerdem großer Wert auf die Validität der Effizienzmaße gelegt. Trotz offenbleibender methodischer Fragen sollte man nicht vergessen, daß es auf die **relativen** und nicht auf die absoluten Werte von unabhängigen und abhängigen Variablen ankommt. Außerdem halten wir die Feldstudien wegen ihrer externen Validität für sehr wichtig. Deshalb sehen wir die Studien bis auf die Untersuchung von Lucas durchaus als geeignet an, die Effizienzhypothesen zum Informationsverhalten zu prüfen.

2.4.2.2. Befunde zur Baustein-Erklärung: allenfalls schwache Belege für die anspruchslose Hypothesenvariante

Die **pauschale Hypothese**, daß die Effizienz um so höher ausfällt, je mehr Daten man beschafft, wird durch die Studien allenfalls schwach belegt:

1. Witte findet keinen signifikanten Zusammenhang zwischen dem Anteil der Informations-Nachfrage-Aktivitäten an allen Aktivitäten eines Entscheidungsprozesses und dem Innovationsgrad, einem Expertenurteil, das mehrere Gütekriterien der konzipierten EDV-Lösung berücksichtigt. Die positive Gesamttendenz führt lediglich bei den Extremgruppen zu signifikanten Unterschieden.²⁶⁾
2. Grinyer und Norburn stellen keine signifikanten Beziehungen zwischen der marktrelativierten Unternehmensrendite vor Zins und Steuern und der Anzahl der Informationen fest, die üblicherweise gefordert werden, um eine eingehende Prüfung der Unternehmensstrategie einzuleiten.²⁷⁾
3. Hauschildt stellt ebenfalls keine signifikante Beziehung zwischen der Nachfrage und der von den Top-Managern 10 Jahre später beurteilten Entscheidungsqualität fest. Erst wenn man die Fälle, die wegen "Zielwandels" als ineffizient beurteilt wurden, ausschließt, zeigt sich ein positiver, tendenziell signifikanter Zusammenhang.²⁸⁾
4. Bei Nutt zeigen die beiden Prozeßtypen, die sich auf die Suche nach extern angefertigten Lösungen beschränken ("off-the-shelf planning" und "search planning"), eine leicht überdurchschnittliche bis geringe Entscheidungsqualität.²⁹⁾

5. In der Studie von Lucas ergeben sich nur wenige signifikante Beziehungen, die der Autor unterschiedlich interpretiert. Da man nicht nur an seiner Erhebung, sondern auch an seiner Auswertungstechnik erhebliche Zweifel anmelden kann,³⁰⁾ sollen die Ergebnisse nicht zur Beurteilung der Hypothesen herangezogen werden.

2.4.2.3. Deutliche Hinweise auf die Gültigkeit der Verarbeitungs-Hypothese

Hinweise auf die Gültigkeit der Verarbeitungs-Erklärung geben die Studien von Grinyer/Norburn und Hauschildt:

1. In der Untersuchung von Grinyer und Norburn, zeigt sich für die Informationen, die bei einer Prüfung der Unternehmensstrategie tatsächlich verwendet werden, eine positive, tendenziell signifikante Korrelation zur Effizienz ($r=0,41$).

2. Bei Hauschildt zeigt sich ein signifikant positiv ausgeprägter Einfluß der Alternativenbildung beim Übergang von zwei auf drei Alternativen. Er ist auch dann vorhanden, wenn man die Informationsnachfrage konstant hält.³¹⁾ Wir teilen die Auffassung Hauschildts, daß die Alternativenbildung die Nachfrage stärker beeinflußt als umgekehrt, denn in einer Auswertung zum gleichen Material weist Haidle nach, daß zur beschlossenen Alternative nicht signifikant mehr Informationen nachgefragt werden als zu den verworfenen.³²⁾ Die positive Wirkung der Informationsaktivität hängt nach der Studie von Hauschildt in erster Linie von der Kreativität bei der Konstruktion des Bauplanes und der Sachgerechtigkeit des daraus abgeleiteten Informationsbedürfnisses ab, nicht von der Menge an Bausteinen, die man anfordert oder unaufgefordert geliefert bekommt.

2.4.2.4. Bestätigung der Steuerungs-Hypothese

Empirische Belege für die Gültigkeit der Steuerungs-Erklärung geben die Untersuchungen von Witte sowie Grinyer und Norburn:

1. In der Studie von **Witte** liegt ein **Interaktionsproblem** mit einem Marktpartner vor. Ob eine Informations-Nachfrage die Effizienz steigert, hängt nicht nur vom Nachfrager, sondern auch vom Informations-Lieferanten ab. Daher kommt es auf die **Abstimmung** von beiden Informationsprozessen an. Betrachtet man die Abweichung beider Aktivitäten, dann zeigt sich eine positive, signifikant ausgeprägte Beziehung zum Innovationsgrad. Die Detailanalyse zeigt: In den Fällen, in denen die Versorgung die Nachfrage übersteigt, führt eine höhere Nachfrage zu einer signifikant höheren Effizienz. Falls aber die Nachfrage bereits über der Versorgung liegt, kann sie keine Effizienzsteigerung mehr bewirken.³³⁾ Die fehlende Effizienzwirkung ist in diesem Falle offenkundig auf Versorgungs-Barrieren des Marktpartners zurückzuführen.

Anzumerken ist, daß Witte diesen Befund zum Anlaß nimmt, eine neue Hypothese zum "Informations-Gleichgewicht" von Nachfrage und Versorgung zu formulieren, die an einem anderen Material oder mit anderen Effizienzgrößen zu prüfen ist. Eine derartige konzeptuelle Replikation wird von Gemünden vorgenommen: Er kann zeigen, daß sich auch bei Effizienzmaßen, die sich auf die Implementierung beziehen und nicht mit dem bei Witte verwendeten "Innovationsgrad" korreliert sind, eine besonders gute Lösung ergibt, wenn die Aktivitäten von Verwender und Hersteller nach Ausmaß und Inhalt aufeinander abgestimmt sind.³⁴⁾

2. Empirische Belege für ein "**Informations-Gleichgewicht**" von Informations-Nachfrage und -Versorgung liefert auch die Studie von Grinyer und Norburn. Relativiert man die zur Strategieprüfung verwendeten Informationen an der Anzahl der gelieferten Informationen, so ergibt sich eine Korrelation von 0,51, die auch bei der relativ kleinen Stichprobe von 21 Unternehmen signifikant ist. In Übereinstimmung mit der interaktionstheoretischen Steuerungserklärung steht auch die tendenziell signifikante Korrelation zwischen der Anzahl zweiseitiger **informeller** Kommunikationsprozesse und der Effi-

2.4.2.5. Zusammenfassende Würdigung der Ergebnisse

Zusammenfassend lassen die Befunde folgende Wertung zu:

1. Zwischen verschiedenen Aspekten des Informationsverhaltens und unterschiedlichen Effizienzmaßen komplexer unternehmenspolitischer Entscheidungen bestehen signifikante, praktisch bedeutsame Beziehungen.
2. Für die anspruchslose Version der Baustein-Hypothese, daß die Effizienz bereits mit der Menge an beschafften Daten steigen werde, lassen sich nur schwache Anhaltspunkte finden. Notwendig scheint eine Differenzierung nach Art der beschafften Bausteine und Vorwissen der Nachfrager.
3. Besondere Stützung erfährt die Steuerungs-Erklärung: Weder die Informationsnachfrage noch die Informationslieferung können für sich genommen eine hohe Effizienz bewirken. Entscheidend ist vielmehr die Abstimmung beider Prozesse. Dies unterstreicht die unternehmenspolitische Bedeutung der Controllingfunktion. Sie besteht nach Horvát "in der Koordination von Informationsversorgung und Informationsverwendung im Führungssystem."³⁶⁾
4. Die Verarbeitungs-Erklärung wird durch den bemerkenswerten Einfluß der Alternativenbildung gestützt. Diese wichtige, Kreativität anzeigende Verhaltensvariable, wird in all jenen Experimenten vernachlässigt, in denen man nur eine "Aus"-Wahl zwischen wenigen "gegebenen" Entschlußoptionen besitzt. Sie wird nur beschränkt erfaßt in (einfachen) Unternehmensspielen, in denen man nur wenige Aktions-"Parameter" innerhalb eines vorgegebenen Wertebereiches variieren darf.

Aus diesen Gründen sollte man in Zukunft mehr Feldstudien durchführen, um die Effizienzwirkungen des Informationsverhaltens zu erforschen und die externe Validität der Schlußfolgerungen, die man aus Laborexperimenten zieht, zu überprüfen. Dies gilt auch dann, wenn man wegen der größeren Komplexität der Entscheidungen und der geringeren Situationskontrolle nur

einen Teil des Informationsverhaltens erfassen kann und sich bei der Effizienzgröße Zurechnungsprobleme ergeben.

2.4.3. Ergebnisse der experimentellen Studien mit computergestützten Managementspielen

2.4.3.1. Informationsnachfrage und Effizienz: zahlreiche Belege für einen positiven Zusammenhang

Wenn eine positive Beziehung zwischen Informationsbeschaffung und intersubjektiven Leistungsunterschieden bei der Lösung komplexer betriebswirtschaftlicher Probleme bestehen sollte, dann sollte sie sich in den nun zu referierenden experimentellen Untersuchungen von computergestützten Managementspielen zeigen:

1. Die **Aufgabe** ist komplex und dynamisch: Zahlreiche Aktionsparameter sind pro Spielperiode festzulegen, die Umwelt reagiert und schafft ständig neue Ausgangsbedingungen.
2. Das **Vorwissen** über die Variablenzustände und ihre Zusammenhänge ist bei allen Versuchspersonen gleichartig niedrig ausgeprägt.
3. Die **Spielbedingungen**, die Ressourcenausstattung, Umweltreaktionen und Aktionsparameter, sind bei allen Versuchspersonen gleichartig ausgeprägt. Eine gewisse Ausnahme bilden lediglich die (wenigen) Wettbewerbsspiele, bei denen die Reaktionen menschlicher Konkurrenten variieren.
4. Die **Effizienzvariablen** sind meistens kardinal skaliert. Sie reagieren deutlich auf die Maßnahmen der Entscheidungsträger und berücksichtigen deren Wirkungen recht umfassend.
5. Das **Informationsangebot** enthält in den meisten Studien überwiegend Informationen, die einen großen Beitrag für die Problemlösung liefern können.

6. Die **Informationsbeschaffung** wird vollständig, zuverlässig und meistens sehr fein abgestuft gemessen.

Unser Ausgangsbefund bestätigt diese Erwartungen (vgl. Abbildung 2.21. und Abbildung 2.22., in der die Studien ausführlicher dargestellt sind):³⁷⁾

* 16 von 18 Testergebnissen, die eine Aussage über die Richtung des Zusammenhanges machen, zeigen eine positive Beziehung. Unter der Nullhypothese, daß in der Grundgesamtheit kein Zusammenhang besteht, ist die Wahrscheinlichkeit für ein derartiges Ergebnis kleiner als 0,0001.³⁸⁾ Selbst wenn man fünf weitere Studien hinzunähme, die eine **negative** Beziehung aufweisen würden, könnte man die Nullhypothese immer noch auf einem Signifikanzniveau von 0,0173 verwerfen. Wir können also davon ausgehen, daß ein positiver Zusammenhang besteht.

* Gewichtet man die errechneten Korrelationen mit dem Stichprobenumfang, ergibt sich eine mittlere Korrelation von 0,31 mit einer Standardabweichung von 0,31. Das 95%-Konfidenzintervall reicht von 0,23 bis 0,39.³⁹⁾

Wir stellen demnach einen wiederholt nachgewiesenen positiven Zusammenhang zwischen Informationsnachfrage und Effizienz fest. Die durchschnittliche Korrelation von 0,31 entspricht einer mittleren Varianzerklärung von 9,6 v. H. Dieser Effekt ist nicht zu vernachlässigen. Er besagt folgendes:

* Nur 35 v. H. der Personen mit einer unterdurchschnittlichen Nachfrage erreichen eine überdurchschnittliche Effizienz.

* Aber 65 v. H. der Personen mit einer überdurchschnittlichen Nachfrage erzielen auch eine überdurchschnittliche Effizienz.⁴⁰⁾

Auf der anderen Seite sind die Unterschiede zwischen den Studien recht beachtlich. Dies ist, wie das nach Hedges und Olkin geschätzte Konfidenzintervall belegt, zu einem nicht zu unterschätzenden Teil auf **Zufallseinflüsse** wegen der ziemlich kleinen Stichprobenumfänge zurückzuführen. Es gibt jedoch auch **systematische Einflüsse**, die das Ergebnis erklären können.

Abbildung 2.21.: Befunde zur Baustein-Hypothese (Computergestützte Managementspiele)

Quelle	Informationsbeschaffung/Effizienzgröße	Stichprobenumfang	Korrelation
Bronner/Witte/Wossidlo (Experimente 1972), Tabelle 85	Gesamte Informationsnachfrage/Gesamterfolg	16 Gruppen	0,51
Bronner (Zeitdruck 1983), Tabelle 72b	Nachfrageaktivitäten/Gesamterfolg	12 Gruppen	0,61
Smith (Inquiry Techniques 1975), Rohdaten auf S. 115 und S. 123	Intensität der Systemnutzung/Kosten	17 Personen	0,03
Chorba/New (Information Support 1980), Tabelle 2	Anzahl nachgefragter Informationen/Rangplatz der Gruppe bezgl. Einzahlungsüberschüsse	31 Personen	0,13
Putz-Osterloh (Problemlöseerfolg 1981), U-Werte Text S. 91 für "transparente Bedingung"	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	13 Personen	0,74
Putz-Osterloh (Problemlöseerfolg 1981), U-Werte Text S. 91 für "intransparente Bedingung"	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	35 Personen	0,40
Putz-Osterloh/Lüer (Problemlöseleistungen 1981), Tabelle 2, "transparente Bedingung"	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	35 Personen	0,63
Putz-Osterloh/Lüer (Problemlöseleistungen 1981), Tabelle 2, "intransparente Bedingung"	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	33 Personen	0,70
Hesse (Semantischer Kontext 1982), Tabelle 4, "mit Kontext, mit Graphik"	Nachfrage nach Zuständen von 12 Systemvariablen/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	0,31
Hesse (Semantischer Kontext 1982), Tabelle 4, "ohne Kontext, mit Graphik"	Nachfrage nach Zuständen von 12 Systemvariablen/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	-0,08
Hesse (Semantischer Kontext 1982), Tabelle 4, "mit Kontext, ohne Graphik"	Nachfrage nach Zuständen von 12 Systemvariablen/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	-0,22
Hesse (Semantischer Kontext 1982), Tabelle 4, "ohne Kontext, ohne Graphik"	Nachfrage nach Zuständen von 12 Systemvariablen/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	0,24
Hesse/Spies/Lüer (Motivationale Faktoren 1983), Text S. 412, hohe Betroffenheit, "Pocken"	Anzahl erfragter Variablenzustände/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	0,52
Hesse/Spies/Lüer (Motivationale Faktoren 1983), Text S. 412, geringe Betroffenheit, "Grippe"	Anzahl erfragter Variablenzustände/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	0,16
Putz-Osterloh (Problemlöseleistung 1983), Tabelle 4, Kontrollgruppe	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	30 Personen	0,13
Putz-Osterloh (Problemlöseleistung 1983), Tabelle 4, Reflexionsgruppe	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	30 Personen	0,35
Putz-Osterloh (Problemlöseleistung 1983), Tabelle 4, Rückmeldegruppe	erfragte Variablen/Anzahl Spielperioden mit steigendem Zahlungsmittelbestand	30 Personen	0,43
Putz-Osterloh (Selbstreflexionen 1985), Tabelle 2, Reflexionsgruppe	Anzahl erfragter Variablen/z-Wert, aufsummiert über verschiedene Gütekriterien	30 Personen	0,33

Abbildung 2.22.: Einflüsse des Informationsverhaltens auf die Effizienz (Computergestützte Managementspiele)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Bronner/Witte/Wossidlo (Experimente 1972)	Unternehmensspiel, Produktion, Absatz, Finanzierung, "Topic 1", mind. 27 Aktionsparameter	72 Manager 72 Stud. verteilt auf 16 Spielgruppen	a) Nachfrage nach Informationen aus dem Prospekt b) Gesamte Informationsnachfrage Jede Spielgruppe besteht aus "Stab" und "Vorstand", die schriftlich miteinander kommunizieren. Nur Stab erhält Betriebs- und Finanzbericht, wichtige Informationen des Vertriebsberichtes sind vom Informationszentrum zu erfragen. Exp. Bed.: "Vitalisierung" der Nachfrage	1. Produktqualität am Spielende 2. Gesamterfolg über zehn Spielperioden	a) n.s. (sign. pos. Unterschiede zw. Extremgruppen) b) - kein Befund - a) n.s. (sign. pos. Unterschiede zw. Extremgruppen) b) positiv sign. (Spearman'sche Rangkorr. Gesamterfolg mit Nachfrageniveau 0,51; Re-Analyse von Tabelle 85)
Bronner (Zeitdruck 1973)	Unternehmensspiel, Produktion, Absatz, Finanzierung, "Topic 1", mind. 27 Aktionsparameter	72 Manager 40 Stud. verteilt auf 12 Spielgruppen	a) Problemlösungsaktivität insgesamt b) Nachfrageaktivitäten c) Prozeßsteuerung d) Funktionssteuerung e) Zielsteuerung Jede Spielgruppe besteht aus "Stab" und "Vorstand", die schriftlich miteinander kommunizieren. Nur Stab erhält Betriebs- und Finanzbericht, wichtige Informationen des Vertriebsberichtes sind vom Informationszentrum zu erfragen. Exp. Bed.: Variation des Zeitdrucks in verschiedenen Spielperioden	1. Gesamterfolg über zehn Spielperioden, Vergleich von je 6 Spielgruppen 2. Rangplätze gegenüber drei konkurrierenden Spielgruppen bezügl. Liquidität, Erfolg, Auftragseingang, Absatz, Produktqualität in jeder Periode, Vergleich von je 60 Meßwerten	a) pos. tend. sign. (p=0,052) b) pos. tend. sign. (p=0,054) c) n.s. d) n.s. e) pos. tend. sign (p=0,052) a) pos. sign. b) n.s. c) pos. sign. d) neg. sign. e) pos. sign. (Testdesign: Meßwerte der Stabsgruppen, Vergleich von Teams mit hoher und niedriger Aktivität)
Lucas (Information Systems 1975), "Laboratory Experiment"	Unternehmensspiel, Logistik, Transportmittel und Wege, Lagerhaltung, "Stanford Business Logistics Game", mind. 12 Aktionsparameter	44 Vpn. Teilmenge aus 36 Stud. u. 78 Man.	Ob bestimmte problemrelevante Informationen auf dem Bildschirm abgerufen werden oder nicht, und zwar in Form von Graphiken. Wird unterschieden in a) "problem-finding" b) "problem-solving" Informationen, wobei für "problem-finding information" vorausgehende geringere Effizienzwerte als auslösende Ursache der Informationssuche postuliert werden	erzielte Gewinne pro Spielperiode	a) "problem-finding": in 11 von 13 Fällen geringere Effizienzwerte der Graphik-Nutzer, aber nur in 5 Fällen sign. Unterschiede bei vorheriger Effizienz b) "problem-solving": 13 der 17 sign. pos. Beziehungen ("Nutzer besser") nach Nutzung, 6 von 9 sign. neg. Beziehungen vor Nutzung, keine klare Aussage, wieviele Beziehungen nicht signifikant sind

Abbildung 2.22.: Einflüsse des Informationsverhaltens auf die Effizienz (Computergestützte Managementspiele)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Schroeder/ Benbasat (Uncertainty 1975)	Unternehmens- spiel, Beschaf- fung und Lager- haltung in einem Handelsunterneh- men, 4 Aktions- parameter, ver- gleichsweise einfaches Spiel	51 Stud.	a) Anzahl nachgefragter Be- richte b) Anzahl getroffener Ent- scheidungen exp. Bed.: Variation der Un- sicherheit über Streuung der Nachfrage, außerdem können je vier versch. Arten von 19 versch. Berichten gegen Ge- bühr von einem Bildschirm abgerufen werden, exp. Bed. hat starke Einflüsse auf die Nachfrage, wird aber bei Effizienzbefund offenbar nicht konstant gehalten	Gesamtkosten	a) "n.s." b) "n.s." (nur pauschale Anga- ben im Text S. 564, evtl. sign. Korrela- tionen mit Nachfra- gemenge sollen sehr gering sein, keine Aussagen über Wir- kungen der Art der nachgefragten Be- richte)
Smith (Inquiry Techniques 1975)	Unternehmens- spiel, Beschaf- fung und Produk- tion, 20 Aktions- parameter, die an fünf verschiede- nen Wochentagen unterschiedlich fixiert werden können	17 Stud.	a) Intensität der Systemnut- zung ("average interac- tions", S. 123) b) Anteil fehlerhafter Nut- zungen ("errors as % of interactions", S. 123) Exp. Bed.: Bildschirmausgabe, strukturiert und aggregiert vs. zusätzliche Option, ei- gene Berichte mit Hilfe eines Datenbanksystems zu erzeugen	Kosten ("mean total cost", S. 115)	a) n.s., Spearman'- sche Rangkorre- lation: 0,03 b) n.s., Spearman'- sche Rangkorre- lation: 0,08 (Re-Analyse: Rang- korr. und Signifi- kanztests aufgrund der bei Smith aus- gewiesenen Daten)
Dentler (Problem- löseverhal- ten 1977)	agrарwirtschaft- lich-ökologisches Managementspiel, 9 Eingriffskate- gorien mit mehr- eren Aktionsparam.	30 Stud. vert. auf 10 Grup- pen	a) Anzahl getroffener Maß- nahmen (Tab. 40) b) Vielfältigkeit ("Breite") der Maßnahmen (Tab. 46) Re-Analyse der Rohwerte: beide Maße als Rangwerte verwendet	Rangsumme von sieben Effi- zienzmaßen (Re-Analyse: Maße auf Basis von Tabelle 48 bis 54 ge- bildet)	a) n.s., Spearman'- sche Rangkorre- lation: +0,22 b) n.s., Spearman'- sche Rangkorre- lation: +0,05 (Re-Analyse aufgrund der in Tab. 40, 46, und 48 bis 54 ausge- wiesenen Rohdaten)
Picot/Lange (Synoptische vs. Inkre- mentale Ge- staltung 1979), vgl. auch Picot/ Lange (Wir- kungsanalyse 1978) Befund S. 42 und 43	Unternehmensspiel, Produktion, Ab- satz, Finanzie- rung, "Topic 1", mind. 27 Aktions- parameter	33 Stud. ver- teilt auf 8 Spiel- grup- pen	a) "synoptische" (zielge- steuerte) vs. "inkre- mentale (ressourcen- orientierte) Vorge- hensweise, exp. indu- ziert durch unter- schiedliche Vorgabe b) Anzahl Alternativen, ermittelt aus vorge- gebenen Planungsbögen und Urteilen der Spieleiter, um "Schein"-Alternativen auszusondern	ROI, Eigenka- pitalrendite, Periodenge- winne, Kapi- talumschlag, Umsatzrenta- bilität	a) "Inkrementale" besser, tend. sign. für Gesamt- periode, sign. für Rezessions- phase (bei Ver- nachlässigung eines "Ausrei- bers" keine sign. Unterschiede b) n.s. Befund wenig aussagekräftig, da "im wesentli- chen" nur eine Alternative diskutiert wird

Anmerkung: Befunde von Autoren "explorativ" genannt, wg. kleiner Fallzahl (8 Gruppen) und Fehlen einer prozeßbegleitenden Messung des Informationsverhaltens nur nachrichtlich vermerkt, Studie verweist auf ähnliche Effizienzprüfungen anderer Forscher

Abbildung 2.22.: Einflüsse des Informationsverhaltens auf die Effizienz (Computergestützte Managementspiele)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Chorba/New (Information Support 1980)	Unternehmens- spiel, Absatz, 4 Wettbewerber je Gruppe, 4 Aktionsparame- ter, einfaches Spiel	13 St. 12 Man. 2 Prof. 4 And.	Anzahl nachgefragter Infor- mationen (gebührenpflichtige Lieferung), insges. werden 96 "data-items" angeboten, von denen aber 64 leicht er- rechnet werden können; exp. Bed.: im vorangehenden Spiel unterschiedl. Informations- angebotsbedingungen, die das Informationsverhalten und die Effizienz stark be- einflussen	Rangplatz pro Gruppe bezgl. kumulierter Einzahlungs- überschüsse	n.s., Korrr. Rangplatz mit Nachfrage: 0,13 (Re-Analyse von "Table 2") Befunde leider nicht nach exp. Einflüssen und Art der nachge- fragten Informa- tionen differen- ziert.
Putz- Osterloh (Problem- löseerfolg 1981)	Unternehmens- spiel, Produk- tion und Absatz; exp. Bed. Trans- parenz vs. In- transparenz der Variablenzusam- menhänge des Computermodells, Befunde getrennt nach Versuchsbed. ausgewiesen; System "Schnei- derwerkstatt, 11 Aktionsparameter	48 Stud.	a) erfragte Variablen b) richtig analysierte Be- ziehungen c) falsch analysierte Be- ziehungen d) Analyse erwartungskon- trärer Beziehungen e) richtige Hypothesen f) falsche Hypothesen g) richtige Planungen h) falsche Planungen i) angemessene Maßnahmen j) unangemessene Maßnahmen k) verschiedenartige Maß- nahmen	Anzahl Spielpe- rioden mit stei- gendem Zahlungs- mittelbestand, Vergleich von schlechten und guten Problem- lösern	a) 2 pos. beide signifikant b) 1 pos., 1 neg. beide n.s. c) 2 neg. beide signifikant d) pos. sign. bei intransp. Bed., kein Befund zur transp. Bedg. e) beide n.s., ohne Richtungsangabe f) 2 neg. sign. g) 1 pos. 1 neg. beide n.s. h) kein Befund i) + j) angemessene Maßnahmen pos., unangem. negativ, illustr. Befunde k) 1 n.s. 1 pos. signifikant
Putz-Oster- loh/Lüer (Problem- löselei- stungen 1981)	Unternehmens- spiel, Produk- tion und Absatz; exp. Bed. Trans- parenz vs. In- transparenz der Variablenzusam- menhänge des Computermodells, Befunde getrennt nach Versuchsbed. ausgewiesen; System "Schnei- derwerkstatt, 11 Aktionsparameter	68 Stud.	a) erfragte Variablen b) richtig analysierte Be- ziehungen c) falsch analysierte Be- ziehungen d) Analyse erwartungskon- trärer Beziehungen e) richtige Hypothesen f) falsche Hypothesen g) richtige Planungen h) falsche Planungen i) angemessene Maßnahmen j) unangemessene Maßnahmen k) verschiedenartige Maß- nahmen (Alle Aktivitäten während des gesamten Prozesses ge- messen, übrige Befunde hier nicht ausgewertet)	Anzahl Spielpe- rioden mit stei- gendem Zahlungs- mittelbestand, Vergleich von schlechten und guten Problem- lösern	a) 2 pos. beide sign. b) 2 pos. beide sign. c) 2 neg. beide n.s. d) kein Befund e) 2 positiv: 1 n.s. 1 tend. sign. f) 2 neg. 1 n.s. 1 sign. g) 2 positiv: 1 sign. 1 tend. sign. h) 2 neg. beide sign. i) kein Befund j) 2 negativ: 1 n.s. 1 sign. k) 1 pos. sign. (für transp. Bed. kein Befund für gesam- ten Prozeß, mit Maßnahmen im ersten Drittel sign. Rangkorrr.)

Abbildung 2.22.: Einflüsse des Informationsverhaltens auf die Effizienz (Computergestützte Managementspiele)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Hesse (Semantischer Kontext 1982)	Managementspiel, einmal als agrarwirtschaftlich-ökologisches, einmal als rein formales Problem präsentiert, System "Dori", 7 Aktionsparameter	120 Stud.	a) Nachfrage nach Zuständen von 12 Systemvariablen, getrennt nach exp. Bed.: 1. ohne vs. mit Pfeildialog der Variablenbeziehungen 2. ohne vs. mit semantischem Kontext sowie: 3. erfolgreichen und erfolglosen Problemlösungen b) Anzahl Entscheidungen über drei kritische Systemvariable, differenziert nach den vier exp. Bedingungen	z-Wert, aufsummiert über verschiedene Gütekriterien	Durchschn. Korrr. Fragehäufigkeit und Güte bei erfolgreichen/-losen Problemlösungen (Mittelung der Korrr. nach Fischer's z-Transformation, Werte aus Tab. 4). m.Kont,m.Gr. 0,40/ 0,22 o.Kont,m.Gr. 0,34/-0,50 m.Kont,o.Gr.-0,27/-0,16 o.Kont,o.Gr. 0,23/ 0,24 b) Durchschn. Korrr. (nach Fischer's z-Transform. Werte aus Tab. 5 bis 7): m.Kontext,m.Grafik 0,72 o.Kontext,m.Grafik 0,41 m.Kontext,o.Grafik 0,31 o.Kontext,o.Grafik 0,33
Dörner/Kreuzig/Reither/Stäudel (Lohhausen 1983)	Managementspiel, politische und ökonomische Entscheidungen in einer Kleinstadt, "Lohhausen", 17 Gruppen von Aktionsparametern, maximal wurden 58 Entscheidungen in einer Periode getroffen, hochkomplexes System	48 Stud.	a) Anzahl Entscheidungen b) Anzahl Entscheidungen pro Absicht c) Hypothesen aufstellen d) Hypothesen prüfen e) Ursachenanalyse f) deduktive Urteile g) kategoris. Urteile h) andere Urteile i) Zielartikulationen j) Prozeßsteuerung (abstrakte "Vornahmen") (Anmerkung: Befunde aus Kapitel 4.4. nicht ausgewertet, da die Nachfrageakt. "Spezifische Orientierung", "Problemsuche", "Effektkontrolle" nicht reliabel gemessen werden).	Beurteilung von sozialen und ökonomischen "Kernvariablen"	a) positiv sign. b) positiv sign. c) n.s. d) positiv tend. sign. (p=0,056) e) positiv sign. f) positiv sign. g) n.s. h) n.s. i) n.s. j) positiv sign.
Gediga/Schöttke/Tücke (Problemlösen 1983)	einfaches agrarökonomisches Spiel, vier Aktionsparameter	28 Stud.	a) verbalisierte Lösungsschritte und Entscheidungen, Vergleich von zwei Gruppen b) Menge und Angemessenheit der Hypothesen über Systemzusammenhänge, nachträgliche Befragung, Bildung von drei Gruppen (Exp. Bed.: statische vs. dynamische Umwelt)	Bevölkerungsentwicklung, Ackerland, Kornspeicher	a) sign. positiv bei allen drei Effizienzkriterien b) sign. positiv bei allen drei Effizienzkriterien (Kein Einfluß der exp. Bedingung auf die Effizienz).

Abbildung 2.22.: Einflüsse des Informationsverhaltens auf die Effizienz (Computergestützte Managementspiele)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Hesse/Spies/ Lüer (Motiva- tionale Fak- toren 1983)	Leiter der Ge- sundheitsbehör- de, binnen 20 Wochen Zahl der an einer Epide- mie Erkrankten herabsetzen, System vermutl. identisch mit "Dori", d. h. 7 Aktionsparameter	60 Stud.	a) Anzahl erfragter Varia- blenzustände b) Anzahl erfragter Varia- blenzustände 5 kriti- scher Systemvariablen Exp. Bed.: niedrige/hohe persönl. Betroffenheit, erzeugt durch Vorgabe Grippe- vs. Pockenepidemie	z-Wert, auf- summiert über ver- schiedene Gütekrite- rien	Durchschn. Kor- r. Fragehäufigkeit mit Effizienz: a) sign. positiv Pocken: 0,52 Grippe: 0,16 b) sign. positiv Pocken: 0,58 Grippe: 0,24 (Unterschiede zw. Korr. Pocken vs. Grippe sign.).
Putz-Oster- loh (Pro- blemlöse- leistung 1983)	Unternehmens- spiel, Produk- tion und Absatz; exp. Bed.: Kon- trollgruppe vs. Erfolgsrückmeldg. vs. Aufforderung eigenes Verhalten zu reflektieren Befunde getrennt nach 1. vs. 2.+3. Drittel sowie exp. Bed. ausgewiesen, System "Schneider- werkstatt", 11 Aktionsparameter	90 Stud.	a) erfragte Variablen b) richtig analysierte Be- ziehungen c) falsch analysierte Be- ziehungen d) Analyse erwartungskon- trärer Beziehungen e) richtige Hypothesen f) falsche Hypothesen g) richtige Planungen h) falsche Planungen i) angemessene Maßnahmen j) unangemessene Maßnahmen k) verschiedenartige Maß- nahmen (Alle Aktivitäten im 1. Drittel gemessen)	Anzahl Spielpe- rioden mit stei- gendem Zahlungs- mittelbestand, Korrelations- koeffizienten	a) 6 pos. davon 3 sign. b) 6 pos. davon 4 sign. c) kein Befund d) kein Befund e) 1 negativ, 5 pos. davon 1 sign. f) 6 neg. davon 4 sign. g) 6 pos., 0 sign. h) 6 neg. davon 3 sign. i) 6 pos. davon 6 sign. j) 1 positiv, 5 neg. davon 3 sign. k) kein Befund
Putz-Oster- loh (Selbst- reflexionen 1985)	Managementspiel, ökologische und wirtschaftliche Entscheidungen; exp. Bed.: Ver- suchsgruppe mit Selbstreflexion u. Kontrollgrup- pe, Befunde ge- trennt ausgewie- sen, System "Moro", 22 Ak- tionsparameter	50 Stud.	a) erfragte Variablen b) richtig analysierte Beziehungen c) falsch analysierte Beziehungen d) Analyse erwartungs- konträrer Beziehungen e) richtige Hypothesen f) falsche Hypothesen g) richtige Planungen h) falsche Planungen i) angemessene Maßnahmen j) unangemessene Maß- nahmen k) verschiedenartige Maßnahmen (Alle Aktivitäten im 1. Drittel gemessen)	z-Wert auf- summiert über verschiedene Gütekriterien (Viehbestände, Bevölkerung, Hungertote, Weidefläche, Grundwasser etc.)	a) 1 pos. sign. 1 n.s. b) 1 pos. sign. 1 n.s. c) 2 n.s. d) kein Befund e) 1 pos. sign. 1 n.s. f) 2 n.s. g) 2 n.s. h) 1 neg. sign. 1 n.s. i) 2 n.s. j) 2 n.s. k) 2 n.s.

So ist bezüglich der **Untersuchungen mit geringen Korrelationen** folgendes festzustellen:

- * Die Untersuchung von **Smith** ist die einzige der hier ausgewerteten, in der ein direkter **Dialog mit dem Computer** stattfindet. In allen übrigen Studien agiert der Versuchsleiter oder ein Informationszentrum als "human interface". Smith weist darauf hin, daß eine **hohe Systemnutzung** auch bei solchen Versuchspersonen zu beobachten ist, die **Schwierigkeiten im Umgang mit dem System** haben.⁴¹⁾ In diesem Fall werden positive Effekte der problembezogenen Nachfrage durch negative Wirkungen der systembezogenen Nachfrage, die der Überwindung von Interaktionsbarrieren dient, überlagert.⁴²⁾
- * In den Studien von Chorba/New, Hesse und Hesse/Spieß/Lüer werden den Versuchspersonen vergleichsweise **einfache Managementspiele** vorgegeben, in denen nur vier bis sieben Aktionsparameter verändert werden können.⁴³⁾
- * Bei der Untersuchung von Chorba und New kommt hinzu, daß das gebührenpflichtige Informationsangebot zahlreiche Informationen enthält, die sich aus den übrigen leicht errechnen lassen. Eine Beschaffung **redundanter Informationen** verstößt gegen die Prämissen der anspruchsvolleren Variante der Baustein-Hypothese.⁴⁴⁾
- * Im Experiment von Hesse/Spieß/Lüer ergibt sich ein signifikant engerer Zusammenhang zwischen Informationsnachfrage und Effizienz in der Versuchsbedingung "Pockenepidemie". Die Autoren führen dies darauf zurück, daß die Bekämpfung einer Pockenepidemie eine höhere subjektive Betroffenheit als die einer Grippeepidemie auslöst. Da unter beiden Bedingungen ungefähr gleich viele Informationen beschafft werden, kann der höhere Zusammenhang bei der Bedingung "Pockenepidemie" auf einer besseren Ausnutzung der Daten, d. h. auf einer Effizienzwirkung der **Informationsverarbeitung** beruhen.

- * In derselben Studie zeigt sich eine engere Beziehung zwischen der Nachfrage nach "kritischen" Systemvariablen und der Effizienz als zwischen der Gesamtnachfrage und der Effizienz. Dies spricht für die zweite Variante der Baustein-Hypothese, die an der **Art** der nachgefragten Informationen anknüpft.

Fazit: Unter den günstigen Bedingungen, die in computergestützten Managementspielen realisiert werden, kann ein nicht zu vernachlässigender Zusammenhang von Informationsnachfrage und Effizienz nachgewiesen werden. Erste Befunddifferenzierungen weisen darauf hin, daß es auch auf die Art der beschafften Bausteine und deren Ausnutzung ankommt. Dies wirft die Frage auf, ob die Überlegenheit erfolgreicher Problemlöser nicht auch auf Vorteilen bei **anderen** Informationsaktivitäten beruht. Zur Beantwortung dieser Frage wollen wir uns nun den Studien zuwenden, die Effizienzwirkungen derartiger Aktivitäten prüfen.

2.4.3.2. Hypothesenbildung, Analysen und Planungen: differenzierende Belege für die Erkenntnis- und Verarbeitungs-Erklärung

Zur Frage, **wie** erfolgreiche Personen Probleme lösen, kann auf mehrere Studien von Putz-Osterloh zurückgegriffen werden. Sie hat verschiedene Experimente mit dem gleichen Kategorienschema ausgewertet und dabei bestimmte Randbedingungen experimentell variiert, so daß wir die Beziehungen zwischen Informationsverhalten und Effizienz über verschiedene Situationen hinweg betrachten können.⁴⁵⁾

Neben den bereits behandelten Fragen nach Variablenzuständen unterscheidet sie Hypothesen, Hypothesenprüfungen, Analysen, Planungen und Maßnahmen:

1. **Hypothesen** sind "Annahmen über Variablen oder Beziehungen zwischen ihnen, über die noch keine Daten vorliegen".⁴⁶⁾
2. **Hypothesenprüfungen** werden definiert als "Vergleiche zwischen Annahmen bzw. Erwartungen und vorliegenden Daten".⁴⁷⁾ In der betriebswirtschaftlichen Terminologie handelt es sich um eine Art "Soll-Ist-Vergleich".

3. Unter **Analysen** versteht die Autorin "Feststellungen von Variablenzuständen oder Angaben über Variablenbeziehungen, die aus vorliegenden Daten abgeleitet werden".⁴⁸⁾
4. **Planungen** bezeichnen "Absichtserklärungen über mögliche Maßnahmen und deren erwartete Effekte".⁴⁹⁾
5. **Maßnahmen** (synonym: Entscheidungen) sind Entschlüsse zur Festlegung von Aktionsparametern.

Diese Aktivitäten werden – soweit möglich – in "angemessene" und "unangemessene" differenziert, und zwar nach dem Kriterium, "ob sie der Systemrealität entsprachen, also richtig waren, oder nicht."⁵⁰⁾

Abbildung 2.23. faßt die Effizienzwirkungen dieser Problemlösungsaktivitäten zusammen. Zum Vergleich sind die bereits referierten Befunde für die Informationsnachfrage aufgeführt.

Die Auszählung läßt folgende Tendenzen erkennen:⁵¹⁾

1. **Falsche Hypothesen und unangemessene Planungen behindern die Problemlösung in starkem Maße.** Zutreffende Hypothesen und angemessene Planungen zeigen wesentlich schwächere Zusammenhänge mit der Effizienz.
2. **Hypothesenprüfungen und richtige Analysen der Variablenbeziehungen steigern die Effizienz.** Falsche Analysen führen jedoch zu keiner Verringerung.
3. **System-angemessene Maßnahmen steigern, unangemessene senken die Effizienz.**

Abbildung 2.23.: Beziehungen zwischen Informationsverhalten und Effizienz
(Re-Analyse verschiedener Untersuchungen von Putz-Osterloh und Lürer)

	Positive Beziehungen			Negative Beziehungen			Ohne Rich- tungs- angabe
	signi- fikant	tend. sign.	nicht sign.	signi- fikant	tend. sign.	nicht sign.	
Nachfrage	8		3				1
Hypothesen richtige falsche	2	1	5	7		1 3	3 2
Hypothesen- prüfungen	3	1	3				
Analysen richtige falsche	7		3	2		1 2	1 2
Planungen richtige falsche	1	2	8	6		1 3	1
Maßnahmen richtige falsche	6		1	4		3	2 2

Gestützt werden diese Befunde zur Erkenntnis- und Verarbeitungs-Erklärung durch Ergebnisse anderer Untersuchungen (vgl. nochmals Abbildung 2.22.):

- * In der Lohhausen-Studie stellen Dörner und Mitautoren fest, daß erfolgreiche Problemlöser mehr Entscheidungen treffen und dabei komplexere Ziele verfolgen.⁵²⁾ Sie prüfen mehr Hypothesen und stellen doppelt so häufig die Frage nach dem "Warum". Ihre Analysen beruhen wesentlich häufiger auf deduktiv abgeleiteten Schlußfolgerungen.⁵³⁾
- * Signifikant positive Einflüsse der Entscheidungsfreudigkeit werden bei Hesse und Gediga/Schöttke/Tücke berichtet. Die Re-Analyse der Daten von Dentler zeigt eine Rangkorrelation von 0,22. Sie ist wegen des geringen Stichprobenumfanges (10 Spielgruppen) nicht signifikant.
- * Differenzierende Analysen bei Hesse und die Befunde in den verschiedenen Studien von Putz-Osterloh und Lürer zeigen allerdings, daß es weniger auf die Menge als vielmehr auf die Art und Dosierung der Maßnahmen ankommt.⁵⁴⁾ Maßgeblich hierfür sind angemessene Hypothesen über die Wirkung der Maßnahmen.

Insgesamt gesehen lassen die Befunde eine **verhängnisvolle Verhaltenskette** der erfolglosen Problemlöser erkennen: **Mangelhaftes Vorwissen führt zu falschen Hypothesen, an denen man festhält. Statt sie zu überprüfen, Daten nachzufragen und Beziehungen zu analysieren, um zu angemessenen Hypothesen zu kommen, werden falsche Planungen aufgestellt und unangemessene Entscheidungen getroffen.**

Wie kann man diese Kette durchbrechen? Eine Möglichkeit besteht darin, die Problemlöser über die "wahren" Beziehungen zwischen den Variablen aufzuklären. Kessler und Ashton bezeichnen diese Art von Aufklärung als "cognitive feedback".⁵⁵⁾ Es soll Erkenntnis stiften, im Gegensatz zum häufig dysfunktional wirkenden "outcome feedback", das nur informiert, ob eine Klassifikation richtig oder falsch gewesen ist.

Aber: Welches "cognitive feedback" soll man geben, wenn der Experimentator die "wahren Beziehungen" nicht kennt und der Entscheidungsträger vor einem **neuen** Problem steht? Wir meinen, daß der Schlüssel in der Änderung des Informationsverhaltens liegt.⁵⁶⁾ Man sollte die Problemlöser dazu anhalten, ihr eigenes Verhalten kritisch zu reflektieren. Hypothesen sind nicht als "gegeben" hinzunehmen, sondern vielmehr systematisch empirisch zu überprüfen. Auch der heuristischen Datenanalyse ist große Bedeutung beizumessen: Gelieferte Informationen sind nicht nur mit Hypothesen, sondern auch mit anderen Informationen zu verknüpfen, um neue Fragen aufzuwerfen, weitere Hypothesen zu stimulieren. Geplante Maßnahmen sind gedanklich durchzuspielen und auch auf ihre "Neben"-Wirkungen hin zu überprüfen.⁵⁷⁾ Eine Informations-"Nach"-Frage in diesem Sinne bietet eine gute Chance für eine erfolgreiche Problemlösung, und zwar auch bei einem anderen als dem gerade gelösten.

Wichtig ist, daß man sich nicht mit einer einmaligen Nachfrage begnügt, sondern **immer wieder** nach den **veränderten** Zuständen kritischer Größen fragt. Eine Reihe von Befunden zum **Verlauf** von Problemlösungsprozessen zeigen, daß erfolgreiche Problemlöser ein hohes Aktivitätsniveau beibehalten oder sogar noch steigern, während es bei erfolglosen Entscheidern abnimmt:

- * Dörner und Mitarbeiter ermitteln bei erfolgreichen Problemlösern ein höheres und steigendes Niveau an Entschlüssen, bei erfolglosen stagniert es auf einem geringen Niveau.⁵⁸⁾
- * Putz-Osterloh stellt fest, daß erfolgreiche Versuchspersonen "... immer wieder **neu** nach dem (veränderten) Zustand der Variablen fragten."⁵⁹⁾
- * Bei Hesse/Spieß/Lüer nimmt die Anzahl erfragter Variablenzustände bei den weniger erfolgreichen, mißerfolgsmotivierten Personen im Zeitablauf ab, bei den erfolgsmotivierten nimmt sie zu.⁶⁰⁾
- * Jacoby/Kuß/Mazursky/Troutman stellen in der weiter unten ausführlicher referierten Studie von Finanzanalysten fest, daß erfolgreiche Analytiker ihr hohes Niveau an beschafften Informationen beibehalten, während es

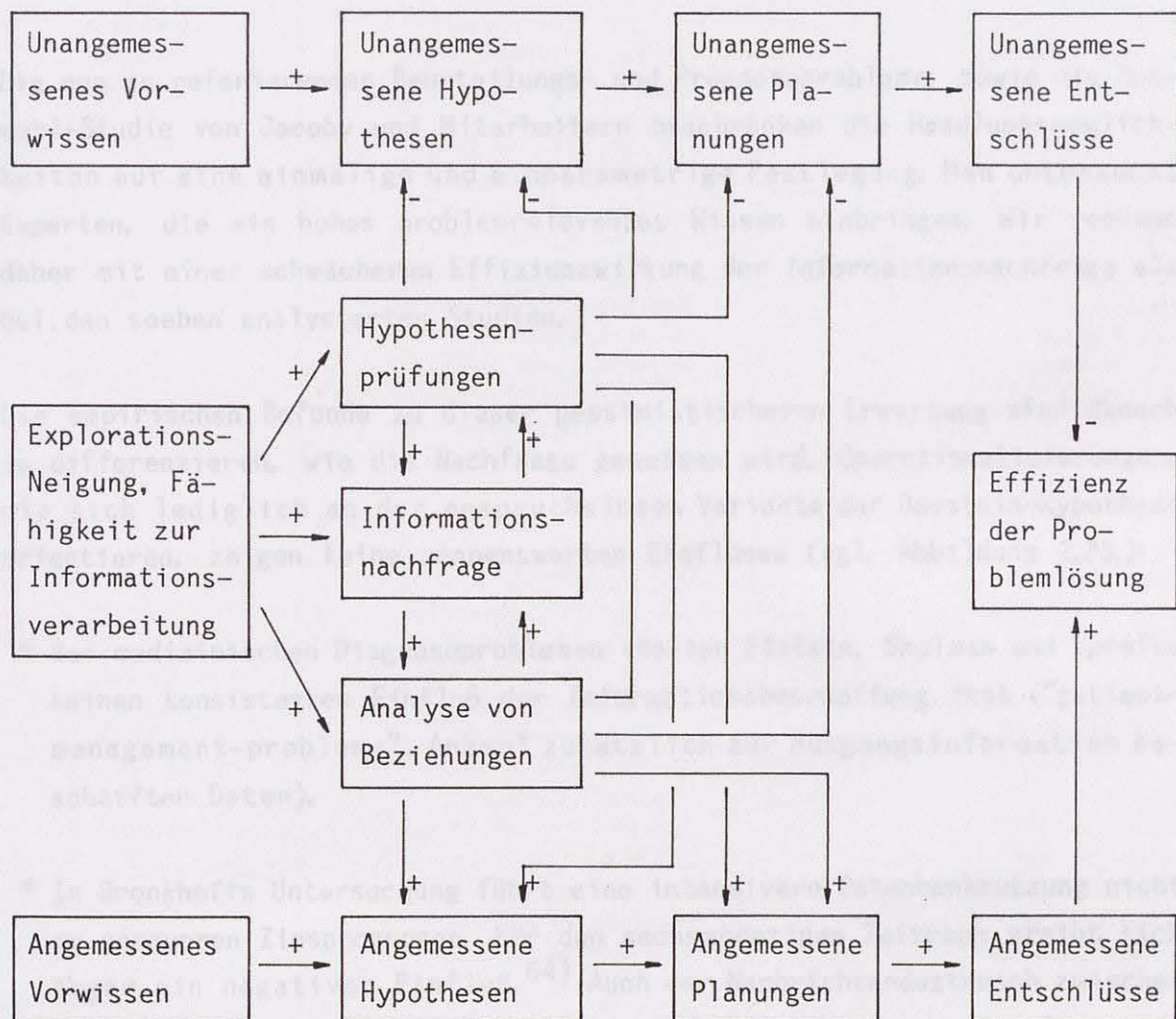
bei erfolglosen signifikant abnimmt.⁶¹⁾

- * In der Untersuchung von Reither unterscheiden sich die erfolgreicherer Expertenteams von Anfängerteams dadurch, daß sie mit einem höheren Entscheidungsniveau starten und auch nach Mißerfolgen beibehalten.⁶²⁾

Schließlich kann man auch Detailergebnisse von Putz-Osterloh heranziehen. In ihrer Veröffentlichung von 1983 hat sie die unzutreffenden Hypothesen nach dem Zeitraum differenziert.⁶³⁾ Man erkennt, daß im ersten Drittel des Unternehmensspiels bei allen drei Experimentalgruppen signifikante Zusammenhänge bestehen: die Korrelationen liegen zwischen $-0,42$ und $-0,53$. Im zweiten und dritten Drittel verschwinden diese Zusammenhänge. Lediglich bei der Gruppe, die im ersten Drittel eine positive Rückmeldung erhielt, und zwar unabhängig von ihrer tatsächlichen Leistung, bleibt eine signifikant negative Korrelation von $-0,44$. Man kann dies als Indiz ansehen, daß fehlerhafte Hypothesen korrigiert und kompensiert werden können, und zwar durch aktive Nachfrage nach Informationen, durch deren kritische Prüfung und durch die Entwicklung von Alternativhypothesen.

So gesehen, ist die oben skizzierte Verhaltenskette kein zwangsläufiger Mechanismus, sondern vielmehr ein mehrstufiger Prozeß, bei dem auf **jeder Stufe** gegengesteuert werden kann, aber nicht muß. Man hätte sich gewünscht, daß derartige Überlegungen in ein pfadanalytisches Modell gemündet hätten, um die Effizienzbeiträge der einzelnen Aktivitätengattungen und der experimentell variierten Rahmenbedingungen deutlicher hervortreten zu lassen. Für ein weiteres Experiment könnte man insbesondere das in Abbildung 2.24. dargestellte Muster pfadanalytisch testen.

Abbildung 2.24. Zusammenwirken von Vorwissen und Informationsaktivität mit der Effizienz der Problemlösung



2.4.4. Ergebnisse der Beurteilungs- und Prognosestudien

2.4.4.1. Befunde zur Baustein-Hypothese: Bestätigung der anspruchsvollen, Falsifikation der anspruchslosen Variante

Die nun zu referierenden Beurteilungs- und Prognoseprobleme, sowie die Auswahl-Studie von Jacoby und Mitarbeitern beschränken die Handlungsmöglichkeiten auf eine einmalige und einparametrische Festlegung. Man untersucht Experten, die ein hohes problemrelevantes Wissen einbringen. Wir rechnen daher mit einer schwächeren Effizienzwirkung der Informationsnachfrage als bei den soeben analysierten Studien.

Die empirischen Befunde zu dieser pessimistischeren Erwartung sind danach zu differenzieren, wie die Nachfrage gemessen wird. Operationalisierungen, die sich lediglich an der anspruchslosen Variante der Baustein-Hypothese orientieren, zeigen keine nennenswerten Einflüsse (vgl. Abbildung 2.25.):

- * Bei medizinischen Diagnoseproblemen stellen Elstein, Shulman und Sprafka keinen konsistenten Einfluß der Informationsbeschaffung fest ("patient-management-problems", Anzahl zusätzlich zur Ausgangsinformation beschaffter Daten).
- * In Brockhoffs Untersuchung führt eine intensivere Datenbanknutzung nicht zu genaueren Zinsprognosen. Für den sechsmonatigen Zeitraum ergibt sich sogar ein negativer Einfluß.⁶⁴⁾ Auch der Nachrichtenaustausch zwischen den Teilnehmern zeigt keinen positiven Zusammenhang mit der Prognosegüte.⁶⁵⁾
- * Shields stellt keinen Einfluß der Beschaffungsmenge auf die Diagnosegenauigkeit der Hauptursache von Unwirtschaftlichkeiten in Kosten- und Leistungsberichten betrieblicher Teileinheiten (sog. "performance reports") fest. Bemerkenswert ist, daß sich auch keine von Shields postulierte umgekehrt-u-förmige Beziehung nachweisen läßt.⁶⁶⁾

Abbildung 2.25.: Einflüsse des Informationsverhaltens auf die Effizienz (Beurteilungs- und Prognoseprobleme)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Elstein/ Shulman/ Sprafka (Problem Solving 1978), "High Fidelity Simulations", Befund S. 95 und S. 97: "average across simulations"	Diagnose von Patienten, Schauspieler simulieren Patienten, außerdem können Laborbefunde nachgefragt werden	24 Ärzte	a) Zeitpunkt der Bildung der ersten Hypothese b) Anzahl gebildeter Hypothesen c) Beschaffte Informationen, bezogen auf die lt. Experten relev. d) Vom Arzt als wichtig erachtete Informationen, bezogen auf die lt. Experten relevanten ("efficiency") e) Anteil der lt. Expertenmeinung angemessen bewerteten Informationen ("accuracy of interpretation")	Ob die nach Expertenmeinung richtigen Diagnosen gestellt werden ("accuracy of outcome")	a) n.s., erkl. Varianz: 1,1 v. H. b) n.s., erkl. Varianz: 2,5 v. H. c) richtig diagnost. Ärzte beschaffen sign. höheren Anteil relev. Inform., erkl. Varianz: 21,4 v. H. d) n.s., erkl. Varianz: 1,6 v. H. e) richtig diagnostizierende Ärzte besitzen sign. höheren Anteil angemessen interpretierter Informationen, erklärte Var.: 20,7 v. H.
Elstein/ Shulman/ Sprafka (Problem Solving 1978), "Patient Management Problems", Befunde S. 128, 132, 138, 141	Diagnose von Patienten, schriftl. Nachfrageoptionen, individueller Ablauf des Informationsprozesses möglich	15 Ärzte	a) Anzahl zusätzlich besch. Informationen b) Anzahl zusätzlich beschaffter Informationen, die nach Expertenurteil problemspezifisch wichtig sind c) Vom Arzt als wichtig erachtete Informationen, bezogen auf die lt. Experten relevanten ("efficiency")	Ob die nach Expertenmeinung richtigen Diagnosen gestellt werden ("accuracy of outcome")	a) bei 1 Problem niedriger, 1 höher, 1 gleich b) bei allen 3 Problemen etwas höher c) bei allen 3 Problemen deutlich höher (Keine Signifikanztests, ein viertes, leichtes Problem von allen Ärzten gelöst)
Brockhoff (Delphi-Prognosen 1979), vgl. auch Brockhoff (Forecasting Quality 1984) mit weiteren Auswertungen desselben Materials	Prognose von Zinssätzen (drei bzw. sechs Monate), in Delphi-Gruppen im Computer-Dialog, zufällige Auswahl von Prognosefragen aus einem Pool von 120 Fragen, je Experte 4 bis 6 Fragen	44 Manager 6 Konjunkturforscher	a) Lektüre von Fachzeitschriften ("x ₁₅ ") b) Datenabfrage aus der Datenbank ("R-Nutzung ₂ ") c) Abfrage von Erläuterungen d) Nachrichtenaustausch zwischen den Teilnehmern einer Delphi-Gruppe Die Datenbank enthält 45 Zeitreihen, die in versch. Präsentationsformen und Aufbereitungsgraden abgerufen werden können.	Relative absolute Abweichung des prognostizierten vom realisierten Wert ("F ₁ ", S. 24)	1. 3-Monats-Prognose Regr. Korr. a) -0,19 -0,26* b) -0,24 -0,02 c) 0,30* 0,25* d) 0,26 -0,07 2. 6-Monats-Prognose Regr. Korr. a) 0,15 -0,30 b) 0,41* 0,43* c) -0,21* 0,14 d) 0,54* -0,03 (Werte aus Tab. 3.17 und Tab. 3.19, "*": auf dem 5%-Niveau sign. einseit. Test)

Abbildung 2.25.: Einflüsse des Informationsverhaltens auf die Effizienz (Beurteilungs- und Prognoseprobleme)

Quelle	Aufgabe	Subjekt	Informationsverhalten	Effizienzgröße	Befund
Abdel-khalik/El She-shai (Information Choice 1980)	Beurteilung, ob ein Firmenkunde Zahlungen einstellt ("default on loans"), exp. stark vereinfachte Problemstellung	28 Kreditmanager	gebührenpflichtige Nachfrage von bis zu acht Kennzahlen, auszuwählen aus 16 angebotenen Kennzahlen	Anteil korrekt klassifizierter Unternehmen	Auswahl der Kennzahl beeinflusst Klassifikationsgüte stärker als deren "korrekte" Verarbeitung, deutl. Unterschiede ohne Signifikanztest
Friedman/Neumann (Opportunity Costs 1980) ^N	Beurteilung von Investitionsprojekten, exp. sehr stark vereinfachte Problemstellung	41 WP 39 Stud.	Beschaffung relevanter Zusatzinformationen, max. 2 aus 5 angebotenen: a) mit vs. ohne Opportunitätskosten b) mit vs. ohne direkte, ausgabewirksame Kosten	Ob das unter Berücksichtigung der Opportunitätskosten günstigste Projekt gewählt wird	a) positiv sign. b) positiv sign. (Mit Beschaffung der relevanten Information wird sign. häufiger die kostengünstigere Entscheidung getroffen)
Shields (Supply and Demand 1983)	Analyse von Kosten- und Leistungsberichten betrieblicher Teileinheiten ("performance reports")	12 Manager je 4 Analysen	a) absolute Menge an nachgefragten Informationen b) relative Menge an nachgefragten Informationen bezogen auf exp. untersch. Angebotsmengen: 3 vs. 9 betriebliche Teileinheiten, 6 vs. 13 Beurteilungskriterien pro Bericht	Ob man die nach Expertenmeinung "kritische Ursache" der Unwirtschaftlichkeit herausfindet	a) n.s., erkl. Varianz: 1 v. H. b) n.s., erkl. Varianz: 1 v. H. (Exp. Bed. beeinflussen Nachfrage und Effizienz sign., Befund wird aber nicht nach exp. Bed. differenziert)
Jacoby/KuB/Mazursky/Troutman/ (Effectiveness 1985)	Aktienkauf, simuliert mit historischen Kursen	17 Finanzanalysten	a) Anzahl beschaffter Informationen b) Anteil beschaffter Informationen aus subjektiv relev. Submatrix c) Art der beschafften Inform. (Anteile der am häufigsten besch. Arten) d) Verlauf der Informationsbeschaffung über die vier Spielperioden Informationsangebot: 8 Aktien, je 26 Informationen. Es dürfen max. 104 Inf. von einem Bildschirm per Lichtgriffel abgerufen werden. Gespielt werden vier Perioden. Der Kauf der Vorperiode hat keinen Einfluß.	Kursentwicklung pro Periode, bereinigt um Dividenden	a) positiv sign. (Korr. 0,41) b) positiv, ohne Signifikanztest Vergleich von Extremgruppen c) positiv, ohne Signifikanztest Vergleich von Extremgruppen (Durchschnittsbefunde über alle 4 Perioden) d) bei 5 besten Analysten gleichmäßig hoch, bei 5 schlechtesten Analysten sign. Abnahme im Zeitablauf

N: Studie wird nur nachrichtlich vermerkt, da sehr stark vereinfachte Problemstellung

- * In der Studie von Jacoby/Kuß/Mazursky/Troutman zeigt sich eine signifikant positive Beziehung zwischen Beschaffungsmenge und Kursgewinn, aber die durch einen Ausreißer nach oben verzerrte Produktmomentkorrelation sinkt von 0,41 auf eine auch im einseitigen Test statistisch nicht signifikante Rangkorrelation von 0,23.⁶⁷⁾

Auf der anderen Seite wird die **anspruchsvolle Variante der Baustein-Hypothese** durch eine Reihe empirischer Befunde gestützt:

- * Bei ihren "high fidelity simulations" stellen Elstein und Koautoren fest, daß Ärzte, die bei allen drei Patienten die richtigen Diagnosen stellen, einen signifikant höheren Anteil an problemspezifisch **relevanten** Informationen beschaffen. Bei drei weiteren "patient-management-problems", die mit Hilfe von Papier-und-Bleistift-Verfahren untersucht werden, bestätigt sich dieser Zusammenhang.⁶⁸⁾
- * Brockhoff findet einen positiven, degressiv steigenden Zusammenhang zwischen Informationsbeschaffung und Prognosegüte, wenn man das **Vorwissen** der Experten konstant hält und die Datenbanknutzung um Verständnisprobleme der Mensch-Maschine-Interaktion bereinigt.⁶⁹⁾
- * Abdel-khalik und El Sheshai ermitteln, daß die Auswahl **geeigneter** Kennzahlen einen größeren Einfluß auf die korrekte Klassifikation von Kreditkunden ausübt als deren angemessene Gewichtung im Sinne einer diskriminanzanalytisch ermittelten Formel.
- * Bei Jacoby und Mitarbeitern ergeben sich bei 16 der angebotenen 26 Informations-**Arten** auf dem 1%-Niveau signifikante Unterschiede bezüglich des Anteils an der gesamten Beschaffungsmenge. Die Detailanalyse zeigt, daß erfolglose Analysten relativ häufiger Informationen nachfragen, die eine **negative** Beziehung zur Kursentwicklung aufweisen.⁷⁰⁾

Fazit: Die Untersuchungen bestätigen die anspruchsvolle Variante der Baustein-Hypothese, daß erst die Beschaffung problemrelevanter und subjektiv neuartiger Informationen die Effizienz steigert. Die anspruchslose Hypothesenvariante wird durch die Befunde nicht gestützt.

2.4.3. Zusammenfassende Würdigung

2.4.4.2. Erkenntnis- und Verarbeitung-Hypothese: positive Evidenz
weniger Prüfungen

Empirische Befunde zur Erkenntnis- und Verarbeitungshypothese werden nur in dem anspruchsvollen Forschungsprojekt von Elstein, Shulman und Sprafka vorgelegt.⁷¹⁾ Sie untersuchen sehr detailliert, welche Rolle die Hypothesenbildung und -prüfung bei medizinischen Diagnoseproblemen spielt. Außerdem prüfen sie, welche Effizienzwirkungen von der Angemessenheit der Bewertung der beschafften Informationen ausgehen. Ihr Projekt belegt zwei Kernthesen:

1. Der Erfolg ärztlicher Diagnoseprozesse wird stark beeinflusst von den formulierten **Hypothesen**. Diese werden sehr früh gebildet und steuern den Verlauf des Prozesses. Wer während des Prozesses einmal die zutreffende Hypothese aufstellt, kommt auch mit hoher Wahrscheinlichkeit zur korrekten Diagnose. Wer eine falsche Ausgangshypothese vertritt, besitzt ein großes Risiko, eine falsche Diagnose zu stellen.⁷²⁾
2. Die Beschaffung problemrelevanter Informationen und ihre angemessene Verarbeitung sind zwei praktisch unkorrelierte Quellen des Diagnoseerfolges. Zusammengekommen erklären sie etwa 40 v. H. der Varianz der Diagnosegüte.⁷³⁾ Daher ist es wichtig, **beide** Verhaltensdimensionen zu betrachten. Der Befund ist auch aus dem Grunde bemerkenswert, weil die **Angemessenheit der Bewertung** der Informationen – im Gegensatz zu den meisten anderen Verhaltensvariablen – über die verschiedenen Fälle hinweg eine sehr hohe Stabilität zeigt.⁷⁴⁾ Sie ist außerdem die einzige Variable, die eine signifikante Beziehung zu dem Experten-Status aufweist, den die Ärzte nach Meinung ihrer Kollegen am Arbeitsplatz besitzen.⁷⁵⁾

Das Projekt bestätigt die Notwendigkeit einer theoriegeleiteten, differenzierten und prozeßbegleitenden Messung des Informationsverhaltens. Es weist darauf hin, die verschiedenen Informationsaktivitäten zueinander in Beziehung zu setzen und ihren Wirkungsverbund pfadanalytisch auszuwerten.

2.4.5. Zusammenfassende Würdigung

Im Gegensatz zu unserer Bestandsaufnahme der Variablen des Informationssystems und der kognitiven Merkmale von Problemlösern zeigt sich beim Informationsverhalten für viele Aspekte ein praktisch bedeutsamer Einfluß.

Die Bestandsaufnahme macht jedoch auch deutlich, daß man in den wenigen Studien, die überhaupt das Informationsverhalten zur Effizienz in Beziehung setzen, meist nur die anspruchslose Variante der Baustein-Hypothese prüft. Während diese These unter den günstigen Bedingungen computersimulierter Managementspiele häufiger bestätigt werden kann, zeigen sich in Feldstudien und in Beurteilungs- und Prognosestudien allenfalls schwache Zusammenhänge. Hierfür sind vermutlich weniger die schlechteren Meßverfahren als vielmehr die anders gearteten Randbedingungen verantwortlich zu machen: Die untersuchten Experten können eine Informationssuche durch bereits vorhandenes Wissen substituieren. Sie sind bereits einer hohen Informationsversorgung ausgesetzt und haben eher das Problem, aus diesen die richtigen auszuwählen und angemessen zu bewerten.

Daß die Art der beschafften Informationen und deren Verarbeitung eine zentrale Rolle spielen, wird durch alle drei Arten von untersuchten Studien belegt. Die Schlußfolgerung kann daher nur lauten, daß erst eine differenziertere, theoriegeleitete Analyse des Informationsverhaltens brauchbare Effizienzerklärungen liefern kann.

3. TEIL: MESSWERTE DER EFFIZIENZ VON BILANZANALYSEN – THEORETISCHES KONZEPT UND EMPIRISCHE VALIDIERUNG –

3.1. THEORETISCHE GRUNDLEGUNG

3.1.1. Ziele der Effizienzmessung

In diesem Teil der Arbeit wird unser Konzept zur Messung der Effizienz von Bilanzanalysen entwickelt und empirisch validiert. Die zentrale Effizienzdimension des Meßkonzeptes, die "Treffsicherheit der Lagebeurteilung", wird im anschließenden vierten Teil zur Prüfung der dort abgeleiteten Effizienzhypothesen herangezogen.

Das Konzept dient nicht nur der Entwicklung von Meßwerten für die empirische Prüfung. Es soll auch eine wichtige Lücke schließen: Die **vorliegenden Ansätze** zur Bestimmung der Effizienz von medizinischen, sozialpsychologischen und betriebswirtschaftlichen Beurteilungsproblemen gehen, wie weiter unten ausführlicher dargestellt wird, von einer stark **vereinfachten Problemstellung** aus.¹⁾ Die betrachtete Aufgabe besteht im allgemeinen darin, wenige vorgegebene Informationen so zu gewichten, daß ein angemessenes klassifikatorisches Gesamturteil gefällt wird. Durch diesen Aufgabenschnitt wird aber gerade die Problematik von Beurteilungsproblemen, die darin besteht, viele Informationen aktiv zu suchen und aus ihnen eine differenzierte Beurteilung abzuleiten, die auch bewußt auf offene Fragen hinweist, verschüttet. Wir sind der Auffassung, daß unser erweitertes Problemverständnis nicht nur bei externen Bilanzanalysen in Kreditentscheidungen, sondern auch bei internen Analysen, die sich auf numerische und verbale Informationen aus der Marktforschung, dem Controlling oder anderen Bereichen stützen, Gültigkeit besitzt. Darüber hinaus vermuten wir, daß man das hier entwickelte Grundkonzept auch zur Bestimmung der Qualität von anderen als rein betriebswirtschaftlichen Beurteilungs- und Analyseprozessen verwenden kann. Trotzdem möchten wir uns darauf beschränken, die Effizienz des bilanzanalytischen Urteilsprozesses zu messen, denn nur für diese Anwendung werden wir eine empirische Validierung vorlegen.

Bilanzanalysen sind komplexe Beurteilungsprobleme, bei denen eine Vielzahl von Informationen zu erfassen, zu verdichten und zu einem Gesamturteil zu integrieren sind. Sie sind zeitaufwendige, ressourcenverbrauchende Prozesse. Von der Qualität der Analyse werden Verlauf und Ergebnis des nachfolgenden Entscheidungsprozesses maßgeblich beeinflusst. Um die Wirtschaftlichkeit der Analyse zu verbessern, sind erhebliche Anstrengungen unternommen worden. Durch maschinelle Erfassung, Verknüpfung und Präsentation der Daten sollen die zahlreichen Routinetätigkeiten des Beurteilungsprozesses rationalisiert werden, durch die Entwicklung von Kennzahlensystemen,²⁾ fragegeleiteten Analysekonzepten³⁾ und geschlossenen Kapitalflußrechnungen⁴⁾ soll die Analysequalität erhöht werden.

Wir bestreiten nicht, daß diese Anstrengungen sinnvoll sind, es irritiert uns aber, daß auf der einen Seite so viele unterschiedliche und sich teilweise widersprechende Vorschläge unterbreitet werden, aber auf der anderen Seite nur so wenige empirische Studien vorgelegt werden, in denen die behaupteten Effizienzwirkungen systematisch und großzählig empirisch geprüft werden. Man kann sich nicht darauf beschränken, die theoretische Angemessenheit von Kennzahleninhalten und Gliederungen von Rechenwerken zu erörtern, man muß auch die Frage stellen, wie diese Informationsangebote auf den Benutzer **wirken**. So mag es sicherlich gute Gründe geben, für bestimmte Entscheidungen Cash-Flow- statt "Gewinn"-Kennzahlen zu empfehlen.⁵⁾ Wenn der ungeübte Benutzer solche Kennzahlen aber für andere als die vorgesehenen Zwecke verwendet oder durch die Berechnungsmethode zu unangemessenen Schlüssen verleitet wird, weil er z. B. glaubt, in einer Verlustsituation durch höhere Abschreibungen mehr "Liquidität" erzielen zu können, dann kann die angestrebte Wirkung leicht verfehlt werden.

Ähnliches dürfte für den zunehmenden Einsatz von Computern gelten.⁶⁾ Zweifelsohne können **Computer** Kennzahlen schneller und zuverlässiger berechnen. Ohne Mühe können sie riesige Kennzahlenlisten ausdrucken oder im Mensch-Maschine-Dialog eine praktisch unbegrenzte Menge berechenbarer Daten bereitstellen. Angesichts der sehr begrenzten menschlichen Verarbeitungskapazität muß man sich aber fragen, ob der Benutzer aus der angebotenen Vielfalt auch die für sein Problem angemessenen Kennzahlen herausgreift. Solche Überlegungen gelten nicht nur für die von uns paradigmatisch untersuchte

externe Bilanzanalyse, sondern auch für Rechenwerke zur Steuerung interner Planungs- und Kontrollprozesse⁷⁾ oder für Frühwarnsysteme.⁸⁾

Aus diesem Grunde reicht es nicht aus, bessere Informations-**Möglichkeiten** anzubieten. Man muß auch empirisch überprüfen, ob und wie sie genutzt werden und welche Effizienzwirkungen sie haben. Unser Ziel ist es, für eine derartige **Effizienzkontrolle** ein geeignetes System von aufeinander abgestimmten Meßwerten anzubieten. Mit diesem System sollen beispielsweise folgende Fragen beantwortet werden können:

- * Welchen Einfluß üben inhaltlich unterschiedlich definierte **Kennzahlen-systeme** aus? Sind komplexe, tiefgegliederte Systeme einfachen, leicht überschaubaren überlegen?
- * Wie wirken unterschiedliche **Präsentationsformen** von Jahresabschlußinformationen?
- * Welchen Beitrag liefern unterschiedliche Varianten der Gestaltung der **Mensch-Maschine-Kommunikation**?
- * Kann man die Qualität der Bilanzanalyse durch **Schulungsmaßnahmen** verbessern?
- * Sind bestimmte **Persönlichkeitsmerkmale** verantwortlich für die Effizienz von Bilanzanalysen? Kommt die Bilanzanalyse bestimmten "Kognitiven Stilen" besonders entgegen?

Um solche Fragen zu beantworten, soll das **menschliche Informationsverhalten** in den Mittelpunkt gestellt werden. Unsere Basishypothese lautet, daß alle Maßnahmen letztlich nur dann die Effizienz steigern können, wenn sie die menschliche Urteilsbildung verbessern.

Die Effizienzwirkungen einer bestimmten Maßnahme sind jedoch sehr komplex. Sie schlagen sich in sehr unterschiedlichen Teilwirkungen nieder, verlaufen über viele Stufen des Analyseprozesses und weisen bei verschiedenen Menschen und Aufgaben ganz unterschiedliche Ausprägungen auf. Um die Komple-

xität der Wirkungsmöglichkeiten angemessen erforschen zu können, sind mehrere **Voraussetzungen** zu schaffen:

1. Eine **standardisierte, experimentell kontrollierte Aufgabe**, anhand derer die Wirkungsanalyse vollzogen wird.
2. Ein **differenziertes Diagnosesystem von Effizienzindikatoren**, um die verschiedenartigen Wirkungen zu erfassen und ihre Interdependenzen analysieren zu können.
3. **Viele Messungen** dieser Indikatoren, um die systematischen von den zufälligen Einflüssen trennen und durchschnittliche Wirkungen innerhalb vertretbarer Konfidenzintervalle quantifizieren zu können.

Gegenstand dieses Kapitels ist die zweite Aufgabe: die Konzeption und empirische Validierung eines Systems von Effizienzindikatoren der Bilanzanalyse. Es soll differenzierte Hinweise liefern, **warum** bestimmte Maßnahmen die Effizienz steigern und bei anderen die angestrebten Effizienzwirkungen ausbleiben.

Bilanzanalysen sind komplexe Informationsprozesse, die bestimmte Anforderungen stellen. Die menschlichen Analytiker können ihnen nur teilweise gerecht werden, weil ihre Informationsaufnahme und -verarbeitungskapazitäten begrenzt sind. Diese **Insuffizienzen** sollen ermittelt werden. Sie sind jedoch keine Konstanten, sondern mehrdimensionale, intersubjektiv unterschiedlich ausgeprägte Größen. Wenn man erklären möchte, wie bestimmte Maßnahmen wirken, dann muß man differenziert erforschen, welche **spezifischen** Insuffizienzen bei bestimmten Personen vorliegen und welche besonderen Funktionen Entscheidungshilfen und Schulungsmaßnahmen erfüllen, um die Insuffizienzen zu beheben.

Das Meßkonzept soll daher folgende **Anforderungen** erfüllen:

1. **Validität:** Die Effizienzmessung sollte sich auf Teilleistungen stützen, die man von der Bilanzanalyse erwartet. Bei der Bewertung dieser Leistungen sind Beurteilungsmaßstäbe anzulegen, die in einer nachgewiesenen positiven Beziehung zur Entscheidungsqualität stehen.
2. **Reliabilität:** Bei Einhaltung der vorgegebenen Beurteilungsregeln sollte sich zu jedem Meßzeitpunkt und bei jedem Beurteiler derselbe Meßwert ergeben (Konsistenz und Konkordanz der Messung). Dies verlangt, daß die Beurteilungsregeln verständlich und praktikabel formuliert sind und einen möglichst geringen Ermessensspielraum gewähren.
3. **Diagnostische Eignung:** Das Konzept sollte für unterschiedliche Teilleistungen eigene Indikatoren bereitstellen (Spezifität) und auf einem möglichst hohen Skalenniveau messen. Die Teilmaße sollten auf unterschiedliche Insuffizienzursachen hinweisen und Anknüpfungspunkte für den gezielten Einsatz von Maßnahmen bieten. Außerdem sollten die Indikatoren in einen systematischen Zusammenhang gebracht werden, so daß man eine Gesamtwirkung leicht in vorgelagerte Teilwirkungen zerlegen kann.
4. **Angemessenheit:** Bei der Bildung der Effizienzmeßwerte sind die situativen Bedingungen des Analytikers zu berücksichtigen. Z. B.: Welches Angebot an Informationen, wieviel Zeit steht ihm zur Verfügung? Mit welcher operationalen Vorgabe soll die Bilanzanalyse erfolgen? Die Anpassung des Meßkonzeptes an derartige Faktoren sollte aber nicht kasuistisch, sondern im Rahmen eines generellen Konzeptes erfolgen, um die Vergleichbarkeit von Untersuchungen, die sich nur in wenigen Aspekten unterscheiden, zu gewährleisten.

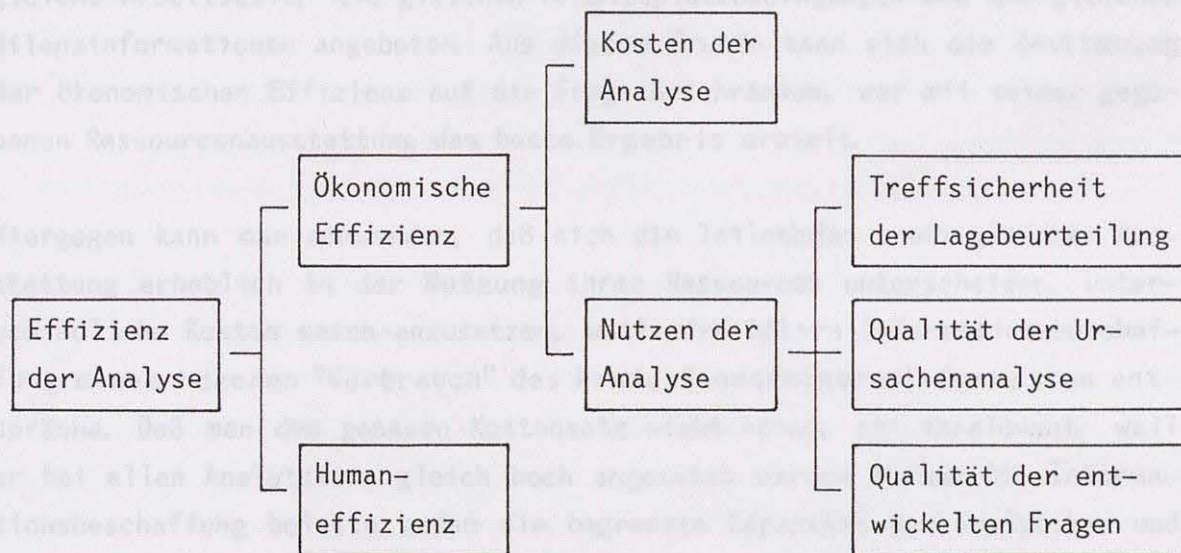
3.1.2. Dimensionen der Effizienzmessung

3.1.2.1. Überblick

Unter **Effizienz** wird ein "zusammenfassendes Urteil über die Erreichung verschiedener Effizienzdimensionen" verstanden.⁹⁾ Als theoretische Basisdimensionen, die einem solchen Urteil zugrundegelegt werden können, sind die

Humaneffizienz und die ökonomische Effizienz anzusehen. Die ökonomische Effizienz wird aufgespalten in Kosten und Nutzen der Analyse. Der Nutzen wird weiter zerlegt in die Dimensionen: Treffsicherheit der Lagebeurteilung, Qualität der Ursachenanalyse und Qualität der entwickelten Fragen (vgl. Abbildung 3.1.).

Abbildung 3.1.: Dimensionen der Effizienz von Bilanzanalysen



3.1.2.2. Humaneffizienz

Als Humaneffizienz soll in unserem Fall die Zufriedenheit der Analytiker mit Verlauf und Ergebnis ihrer Analyse angesehen werden. Die Humaneffizienz interessiert uns vor allem wegen ihrer möglichen Zusammenhänge mit den Effizienzdimensionen des Analysenutzens. Wenn sich zeigen sollte, daß subjektive Mißerfolgserlebnisse für objektiv schlechte Analyseleistungen verantwortlich zu machen sind, dann ist hieraus ein Hinweis auf die Gestaltung des Leistungserlebnisses abzuleiten. Wir vermuten jedoch aufgrund der uneinheitlichen Ergebnisse zur Arbeits- und Verbrauchierzufriedenheit,¹⁰⁾ daß nur eine mäßig positiv ausgeprägte Beziehung zwischen objektiv bewerteter und subjektiv erlebter Leistung besteht, weil sich die Analytiker an sehr

unterschiedlichen Anspruchsniveaus orientieren. Daher wird die Humaneffizienz von uns nur relativ grob durch subjektive Ratings der Versuchspersonen gemessen und nicht weiter in mögliche Ineffizienursachen zerlegt.

3.1.2.3. Kosten der Analyse

Die **Ressourcenausstattung** wird in unseren experimentellen Untersuchungen **konstant** gehalten: Allen Teilnehmern werden die gleichen Arbeitsmittel, die gleiche Arbeitszeit, die gleichen Arbeitsplatzbedingungen und die gleichen Bilanzinformationen angeboten. Aus diesem Grunde kann sich die Bestimmung der ökonomischen Effizienz auf die Frage beschränken, wer mit seiner gegebenen Ressourcenausstattung das beste **Ergebnis** erzielt.

Hiergegen kann man einwenden, daß sich die Teilnehmer trotz gleicher Ausstattung erheblich in der **Nutzung** ihrer Ressourcen unterscheiden. Unterschiedliche Kosten seien anzusetzen, weil eine höhere Informationsbeschaffung einem höheren "Verbrauch" des Produktionsfaktors Information entspräche. Daß man den genauen Kostensatz nicht kenne, sei irrelevant, weil er bei allen Analytikern gleich hoch angesetzt werden könne. Die Informationsbeschaffung belaste zudem die begrenzte Kapazität der Analytiker und könne daher Opportunitätskosten in Form einer verminderten Analysequalität hervorrufen. Deshalb könne man die Menge an beschafften Informationen als Indikator für einen kostenverursachenden "Ressourcenverbrauch" ansehen.¹¹⁾

Gegen die (ausschließliche) Interpretation der Informationsbeschaffung als "Kosten" sprechen folgende Vorbehalte:

1. Wir müssen unterscheiden zwischen der kostenverursachenden Beschaffung eines Informationspakets¹²⁾ und der Nutzung einzelner Informationen aus diesem Paket. Wenn man aus einem bereits erworbenen Informationsbestand einzelne Informationen abrufen, dann entstehen meist nur vernachlässigbare Kosten. Die Information bleibt physisch erhalten und kann auch für andere Benutzer noch einen ökonomischen Wert besitzen. Der **Gebrauch** eines Informationsgüterbestands ist nicht mit dem Ressourcenverbrauch eines Sachgüterbestands gleichzusetzen.¹³⁾ In diesem Punkt ist unser

Experiment sehr realitätsnah, denn in einer realen Kreditentscheidung werden üblicherweise keine Kosten für einzureichende Jahresabschlüsse erstattet. Die Kreditinstitute können nämlich auf ihre gesetzlich geregelte Verpflichtung verweisen, die Bilanzen ihrer potentiellen Kreditnehmer zu prüfen.¹⁴⁾

2. Die vorliegenden empirischen Untersuchungen belegen einen starken Einfluß der Kosten der Informationsbeschaffung auf die Informationsnachfrage.¹⁵⁾ Wir können daher nicht davon ausgehen, daß sich die Versuchspersonen in einer Situation mit gebührenpflichtiger Beschaffung genauso verhalten hätten. Es würde gegen den Grundsatz der Angemessenheit der Effizienzbeurteilung verstoßen, wenn man einen Analytiker als ineffizient bezeichnet, weil er sich in einer gebührenfreien Situation Informationen beschafft, die er sonst nicht beschaffen würde.

3. Eine höhere Informationsbeschaffung **kann** zu höheren Opportunitätskosten in Form eines reduzierten Analysenutzens führen, wenn die verbrauchte Zeit für das Einholen, Aufbereiten und Abspeichern zusätzlich beschaffter Informationen tatsächlich dazu führt, daß die Informationsverarbeitung reduziert wird oder weniger sachgerecht erfolgt. Das Problem besteht jedoch darin, diese funktionalen Abhängigkeiten zu quantifizieren, und zwar bei jedem einzelnen Analytiker, denn man muß mit starken interindividuellen Unterschieden rechnen. Problematisch wäre ein pauschaler Opportunitätskostensatz. Unsere Bestandsaufnahme im zweiten Teil der Arbeit zeigte, daß es **keine** generelle Beziehung zwischen Angebotsmenge und Effizienz gibt, auf die man sich stützen kann, obwohl andere Untersuchungen einen deutlich ausgeprägten positiven Zusammenhang zwischen Angebots- und Beschaffungsmenge belegen.¹⁶⁾

Wegen der ungeklärten Beziehungen zwischen der Informationsbeschaffung und möglichen Opportunitätskosten in Form eines reduzierten Analysenutzens halten wir auch wenig davon, die ökonomische Effizienz in Form eines "Nutzen/Kosten"-Quotienten zu messen, wobei man die "Kosten" durch die Beschaffungsmenge operationalisiert, denn eine derartige Kennzahl ist nur dann aussagekräftig, wenn zwischen beiden Größen eine proportionale Beziehung besteht.¹⁷⁾ Es ist zweckmäßiger, die tatsächlichen Beziehungen zwischen

beiden Größen empirisch zu erforschen, um auf Basis dieses **Wissens** gesicherte Aussagen treffen zu können. Wir warten daher unsere empirische Analyse der Beziehungen zwischen Informationsverhalten und Effizienz ab, die im vierten Teil der Arbeit erfolgt. Nur sie kann die Frage beantworten, ob es überhaupt Überlastungseffekte gibt und wenn ja, ab welcher Nachfragemenge und mit welcher Streubreite sie zu erwarten sind.

3.1.2.4. Nutzen der Analyse

Der Nutzen der Analyse ist die zentrale Dimension unseres Effizienzkonzeptes. Welchen Nutzen die Bilanzanalyse stiftet, hängt von den Aufgaben ab, die sie erfüllen soll. Mit Hauschildt vertreten wir die Auffassung, daß die Bilanzanalyse nur eine Teilaktivität in einem mehrstufigen Beurteilungsprozeß ist. Sie dient nicht dazu, ein endgültiges Urteil zu begründen, sondern sie soll weitere Entscheidungsaktivitäten vorbereiten.¹⁸⁾ Der Nutzen der Bilanzanalyse ist daher an der **Qualität der Entscheidungsvorbereitung** zu messen.

Aus diesem Grunde haben wir uns im Gegensatz zu den meisten Experimenten in der anglo-amerikanischen Literatur nicht darauf beschränkt, die Teilnehmer lediglich um ein Gesamturteil oder eine Prognose einer ausgewählten Kennzahl zu bitten, an denen man die "Effizienz" leicht hätte messen können.¹⁹⁾ Die Versuchspersonen wurden vielmehr aufgefordert, die Rolle eines Kredit-sachbearbeiters zu übernehmen, der einen fiktiven Vorgesetzten auf eine Kreditverhandlung mit einem Firmenkunden vorbereiten soll. Die dazu durchzuführende Bilanzanalyse sollte dem Vorgesetzten eine differenzierte Beurteilung der Unternehmung und eine Liste offener Fragen liefern, deren Antworten er in der nachfolgenden Kreditverhandlung einholen konnte. (Zur Beschreibung der konkreten Aufgabenformulierung siehe Kapitel 4.2.2.).

Zur Bestimmung des komplexen Konstrukts "Nutzen" der Bilanzanalyse ist eine Zerlegung in mehrere Teildimensionen zweckmäßig. Unseres Erachtens wird die Qualität der Entscheidungsvorbereitung von folgenden Leistungen bestimmt:

- a) **Urteilen** zur Bewertung der wirtschaftlichen Lage des Unternehmens.
- b) **Erklärungen** für die Ursachen der festgestellten Lageentwicklung.
- c) **Offenen Fragen** zu vermuteten Ursachen und latenten Gefährdungsmomenten und Chancen, die aus den Bilanzdaten nicht ersichtlich sind.

Zu a) Lagebeurteilung

Die Bilanzanalyse kann sich nicht darauf beschränken, Daten aufzubereiten und Kennzahlen zu berechnen. Sie muß die Kennzahlen auch **beurteilen**, indem sie Vergleiche anstellt und systematische Beziehungen zwischen verschiedenen Kennzahlen herstellt. Die Kennzahlen sollten zweckentsprechend gebildet, die Vergleiche sachgerecht relativiert werden. Außerdem sollten die Urteile auf faktisch und logisch wahren Prämissen beruhen. Diese **instrumentellen Anforderungen der Urteilsbildung** sind eine notwendige Voraussetzung, um den Adressaten vor Irrtümern zu bewahren.

Für einen hohen Analysenutzen kommt es aber nicht nur darauf an, wie beurteilt wird, sondern auch, **was** beurteilt wird, d. h. welche **Relevanz** die beurteilten Sachverhalte besitzen. Unser weiter unten ausführlich dargestelltes Meßkonzept stützt sich auf drei Gewichtungskomponenten: die Empfehlungen der bilanzanalytischen Literatur, die statistisch ermittelten Unterschiede von erfolglosen und erfolgreichen Unternehmen und die fallspezifischen Ausprägungen repräsentativer Kennzahlen der Urteilsbereiche. **Der sachgerecht urteilende, erfahrene, belesene und aufmerksame Analytiker bildet somit die Leitfigur unserer Effizienzdimension "Treffsicherheit der Lagebeurteilung".**

Zu b) Ursachenforschung

Eine gute Bilanzanalyse kann sich nicht darauf beschränken, wichtige Tatbestände sachlich richtig zu berichten. Sie muß die Sachverhalte auch in einen Kontext stellen, durch vorgelagerte Ursachen **erklären** und auf latente Risiken und Chancen hinweisen – soweit das Bilanzmaterial dies zuläßt. Eine gute Ursachenanalyse ist die Voraussetzung für einen präventiven Kredit-

schutz, der sich das Ziel setzt, Geschäftsbeziehungen zu erhalten und auszubauen, statt die Kreditnehmer lediglich nach "guten" und "schlechten" Risiken zu sortieren.²⁰⁾ Ursachen gehen den Wirkungen voraus. Nur wenn man die Ursachen einer krisenhaften Entwicklung erkennt, kann man rechtzeitig und gezielt Gegenmaßnahmen ergreifen oder sie den Firmenkunden empfehlen.

Wie kann man die Qualität einer Ursachenanalyse messen? Naheliegender wäre es, für jeden experimentellen Fall einen Ursachenkatalog aufzustellen und zu überprüfen, ob der Analytiker diese Ursachen zutreffend diagnostiziert.²¹⁾ Gegen eine solche Vorgehensweise sprechen zwei Gründe:

1. Von einer gesicherten Theorie der Unternehmenskrise ist man weit entfernt. Erst allmählich setzt sich die Einsicht durch, daß es nicht "die" Krise, sondern sehr unterschiedliche Formen gibt. Die bisher vorgelegten Listen vermuteter Insolvenzursachen²²⁾ und Typen des "Mis-Managements"²³⁾ können allenfalls als Illustrationen typischer Symptome angesehen werden. Es wird weder eine theoretisch begründete Beziehung zwischen spezifischen Ursachen und bestimmten Wirkungen hergestellt, noch der empirische Nachweis erbracht, daß die vermuteten Ursachen den Wirkungen tatsächlich vorausentfallen. Unser Ursachenkatalog wäre demnach unvollständig, kasuistisch und nicht empirisch abgesichert.

2. Es ist außerordentlich schwierig zu beurteilen, welche dieser "Ursachen" man aufgrund der angebotenen Bilanzinformationen hätte aufspüren müssen. Die recht umfangreiche empirische Bilanzforschung beschränkt sich mehrheitlich darauf, charakteristische Unternehmensentwicklungen mit Hilfe hochaggregierter Kennzahlen zu prognostizieren.²⁴⁾ Ob und wie sich spezifische Krisenursachen in Bilanzdaten niederschlagen, wird im allgemeinen nicht erforscht.²⁵⁾

Wir bewerten daher nicht das Ergebnis, sondern den **Prozeß** der Ursachenanalyse. Dies hat den Vorteil, daß man sich nicht auf möglicherweise zufällig zutreffende Entdeckungen bei einem Einzelfall verläßt, sondern eine **generalisierbare Fähigkeit**²⁶⁾ erfaßt, die auch bei anderen Fällen zu einer besseren Ursachenfindung führt.

Eine Ursachenanalyse umfaßt folgende Aktivitäten:

1. Lokalisierung von Schwächen und Stärken,
2. Entwicklung von Hypothesen, die das Auftreten von Stärken und Schwächen erklären,
3. Überprüfung dieser Hypothesen mit Hilfe geeigneter Informationsverknüpfungen,
4. Bewertung der Ergebnisse.

Zur Lokalisierung von Schwächen und Stärken kann man Kennzahlen in Komponenten zerlegen und die davon wichtigen weiter verfolgen. So kann man beispielsweise einen Jahresfehlbetrag auf einen Betriebsverlust und diesen auf einen starken Umsatzrückgang und einen hohen und deutlich gestiegenen Materialanteil an der Gesamtleistung zurückführen. Diese routinemäßige Komponentenanalyse²⁷⁾ wollen wir ausdrücklich vernachlässigen, weil sie noch keine kausalen, sondern nur definitorische Beziehungen herstellt und die beurteilten Sachverhalte bereits in die Bewertung der Treffsicherheit der Lagebeurteilung eingeflossen sind. Unser Meßkonzept ist spezifischer: es erfaßt, welche bilanzanalytisch besonders anspruchsvollen Verknüpfungen von Sachverhalten der Analytiker vornimmt, um seine Hypothesen zu überprüfen. Die gemeinten Verknüpfungen zeichnen sich häufig dadurch aus, daß man Sachverhalte, die in unterschiedlichen Rechenwerken abgebildet werden, zueinander in Beziehung setzt und daß mehrstellige Relationen bildet. Nach Bouwman's Studie ist dies ein Merkmal von Experten der Bilanzanalyse.²⁸⁾ Nach den Ergebnissen der neueren psychologischen Problemlöseforschung zeichnen sich erfolgreiche Problemlöser dadurch aus, daß sie Kausalnetze statt Kausalketten bilden,²⁹⁾ daß sie mehr angemessene Hypothesen bilden und diese empirisch überprüfen.³⁰⁾ Die von uns gemessenen Verknüpfungen erfassen somit eine generelle Anforderung erfolgreichen Problemlösens. Da die Verknüpfungen fallspezifisch neu zu generieren sind, setzen sie nicht nur hohen bilanzanalytischen Sachverstand, sondern auch Vorstellungskraft und Kreativität voraus.

Zu (c) Entwicklung von Fragen

Nicht alle Hypothesen, die der Analytiker entwickelt, können mit Hilfe der angebotenen Bilanzdaten überprüft werden. In diesem Fall muß die Ursachenanalyse mit anderen Mitteln fortgesetzt werden: Der Bilanzanalytiker formuliert offene Fragen, die durch nachfolgende Informationsaktivitäten beantwortet werden (sollen). Diese Fragen brauchen sich nicht darauf zu beschränken, Ursachen bereits offenkundig gewordener Schwachstellen aufzudecken. Sie können sich auch auf **latente** Gefährdungsmomente richten.³¹⁾ Offene Fragen sind besonders wichtig, weil sich ursächliche Vorgänge in der Unternehmensumwelt und in den betrieblichen Teilbereichen erst relativ spät und auf einem hoch aggregierten Niveau in der Bilanz niederschlagen. Um die Qualität der entwickelten Fragen zu messen, wird berücksichtigt, ob sie auf faktisch und logisch wahren Prämissen beruhen, nicht aus den angebotenen Bilanzdaten beantwortet werden können, eine hohe Relevanz besitzen und mit welcher Begründung sie artikuliert werden.³²⁾ Damit soll nicht nur erfaßt werden, welchen Nutzen die Fragen dem Adressaten liefern, es soll auch gemessen werden, welche **Problemeinsicht** der Analytiker gewonnen hat. Nach der "Erkenntnis"-Hypothese von Witte führt nämlich bereits eine **unversorgte** Informationsnachfrage zu einer Steigerung der Effizienz einer Entscheidung, weil bereits die Formulierung der Frage eine Problemstrukturierung voraussetzt.³³⁾ Da die Entwicklung von Fragen an den fallspezifischen Besonderheiten anknüpft, liegt auch dieser Effizienzdimension die Leitfigur eines sachverständigen und kreativen Analytikers zugrunde.

3.1.3. Vergleich mit Meßkonzepten anderer Untersuchungen

3.1.3.1. Überblick

Mit diesen Ausführungen sind die Ziele und Inhalte unseres Meßkonzeptes dargelegt. Seine Basisideen sind die Orientierung der Effizienzmessung an der Qualität der Entscheidungsvorbereitung und die differenzierte Diagnose von Insuffizienzen des Analytikers. Zum Vergleich soll nun belegt werden, welche Effizienzmeßwerte in anderen experimentellen Untersuchungen verwendet werden.

Folgende Gruppen von Maßgrößen werden häufiger benutzt:

- a) Ressourcenverbrauch (Entscheidungszeiten),
- b) Subjektive Leistungseinschätzung (Leistungszufriedenheit, Konfidenz),
- c) Ökonomischer Erfolg (Gewinne und Kosten in Unternehmensspielen),
- d) Prognosegüte (Vergleich von prognostiziertem und tatsächlichem Ereignis),
- e) Angemessenheit von Beurteilungen (Vergleich von klassifiziertem und tatsächlichem Ereignis).

Diese Maßgrößen unterscheiden sich von unserem Konzept, weil sie für **andere Zwecke** entwickelt wurden. Es soll daher auch kein Konflikt konstruiert werden, um mögliche "Vor"- oder "Nachteile" hervorzuheben. Wir möchten vielmehr die unterschiedlichen Aufgaben verdeutlichen und Verknüpfungsmöglichkeiten aufzeigen. In den **Übersichten am Ende des Kapitels 3.1.** sind die vorgenommenen Operationalisierungen zusammenfassend dargestellt. Zur Veranschaulichung der untersuchten Informationsprobleme ist auch die jeweils zugrundeliegende Informationsbasis beschrieben.

3.1.3.2. Ressourcenverbrauch

Als Maßgröße des Ressourcenverbrauchs wird im allgemeinen der **Zeitaufwand** zur Lösung einer gestellten Aufgabe verwendet. Abbildung 3.3. zeigt, daß die Entscheidungszeiten sowohl innerhalb als auch zwischen den empirischen Studien erhebliche Streuungen aufweisen. Generell liegen die Zeiten aber meistens unter zwei Stunden, so daß man die ökonomische Bedeutung dieser leicht erhebbaren Effizienzgröße nicht überbewerten sollte. In unseren Experimenten war die verfügbare Zeit (90 Minuten) konstant gehalten und von der großen Mehrheit der Teilnehmer auch ausgeschöpft worden. Der nachträglich erfragte wahrgenommene Zeitdruck steht zu keiner unserer Effizienzgrößen in einer signifikanten Beziehung. Wir werden diese Variable

daher bei den nachfolgenden Betrachtungen vernachlässigen.

3.1.3.3. Subjektive Leistungseinschätzung

Zufriedenheits- und Akzeptanzmaße werden zu wichtigen Effizienzgrößen, wenn Entscheidungshilfen durchgesetzt werden sollen. Diese Anwendungsvoraussetzungen stellen jedoch nur eine notwendige Bedingung für die Wirksamkeit von Entscheidungshilfen dar. Sie entbinden die betriebswirtschaftliche Forschung nicht von der Verpflichtung, empirisch zu prüfen, ob die zur Implementierung vorgeschlagenen Entscheidungsunterstützungssysteme tatsächlich zu einer Verbesserung der Entscheidungsqualität beitragen, denn andernfalls

"... besteht die Gefahr, daß Realisierbarkeit und Akzeptanz eines solchen Systems einseitig in den Vordergrund gerückt werden."³⁴⁾

Wir sind dem Impuls der Studien, die subjektive Leistungseinschätzungen verwenden, jedoch insoweit gefolgt, als wir bereits bei unseren experimentellen Anwendungsbedingungen untersuchen, welche Beziehung zwischen der Entscheidungsqualität und der subjektiven Zufriedenheit besteht.

3.1.3.4. Ökonomischer Erfolg

Maßgrößen des ökonomischen Erfolges, wie z. B. Spielgewinne und Kosten werden, wie Abbildung 3.5. belegt, mit Abstand am häufigsten in experimentellen Effizienzprüfungen herangezogen. Sie sind vermutlich deshalb so beliebt, weil sie keinen Zweifel lassen, daß eine "betriebswirtschaftliche" Problematik untersucht wird. Außerdem kann man Ausbildungs- und Forschungsziele bei Anwendungen computergestützter Managementspiele gut miteinander verbinden.

Wenn ökonomische Maßgrößen verwendet werden, dann wird eine längere Wirkungskette als in unserem Experiment betrachtet. Man kann beobachten, wie sich eine bestimmte Eingabegröße auf eine sehr weit entfernte ökonomische Ergebnisgröße auswirkt. Damit vergrößert sich jedoch auch die **Zurechnungs-**

problematik. Die betrachtete Effizienzwirkung hängt nicht mehr ausschließlich von der Qualität der verbesserten Entscheidungsvorbereitung, sondern auch von der Verwendung der unterbreiteten Vorschläge und den modellierten Umweltreaktionen ab. Wir haben uns daher darauf beschränkt, nur die Entscheidungs-**Vorbereitung** zu betrachten, um diese komplexe Phase differenzierter würdigen zu können. Es liegt jedoch auf der Hand, daß man eine vereinfachte Fassung unseres Instruments in eine experimentell umfassender abgebildete Kreditentscheidung einbringen kann. Andererseits ist zu bedenken, daß in den experimentell simulierten Entscheidungen nicht nur **motivationale Artefakte der Spielsituation** das Ergebnis beeinflussen. Auch die Abbildung der realen Entscheidungssituation wirft erhebliche Validitätsprobleme auf: Was angemessen ist und zu "effizienten" Entscheidungen führt, entscheidet nicht nur der Spieler, sondern auch das im Computer realisierte Modell.³⁵⁾

3.1.3.5. Prognosegüte

Prognosen stellen zweifelsohne wichtige Ergebnisse einer Entscheidungsvorbereitung dar. Es wäre relativ leicht gewesen, auch in unserem experimentellen Design eine Prognoseeffizienz zu bestimmen. Wir hätten lediglich eine geeignete Kennzahl auswählen und deren prognostizierte mit der tatsächlichen Entwicklung vergleichen müssen. Dieser Plan klingt sehr verlockend, löst aber unser Problem nicht: Bilanzanalysen und Prognosen ausgewählter Kennzahlen sind unterschiedliche Aufgaben. Die Prognose konzentriert sich darauf, die ausgewählte Kennzahl so präzise vorherzusagen, wie es die zu treffende Entscheidung erfordert. Die Bilanzanalyse dient der Überprüfung vielfältiger Verhaltenserwartungen und der Aufdeckung fallspezifischer Gefährdungsmomente und Entwicklungsmöglichkeiten, um weitere Informationsaktivitäten vorzubereiten und Anregungen für geeignete Handlungsmöglichkeiten zu geben. Die Anforderungen beider Aufgaben weisen sicherlich gewisse Überschneidungen auf, aber sie werden sich kaum vollständig decken. Wenn man nun zusätzlich zur "normalen" Bilanzanalyse eine spezifische Prognoseaufgabe stellt, dann legt man damit zwar eine weitere Anforderung fest, aber man hat noch nicht geklärt, woran man die Qualität des Teils der Bilanzanalyse messen soll, der nicht der Prognose einer ausgewählten Kenn-

zahl dient. Außerdem ist zu bedenken, welches Gewicht die beiden konkurrierenden Aufgaben erhalten sollen, wie man diese Sollvorstellung durchsetzt und die individuell unterschiedlichen Aufgabenmischungen vergleichbar macht.

3.1.3.6. Angemessenheit von Beurteilungen

Das Effizienzmaß "Angemessenheit der Gesamtbeurteilung" bezieht sich wie die Prognosegüte auf den Vergleich von tatsächlicher und behaupteter Entwicklung. Im Unterschied zur Prognose wird jedoch keine kardinale Aussage über die zukünftige Entwicklung einer ausgewählten quantitativen Größe gemacht. Man muß vielmehr ein kategoriales Urteil über ein zukünftiges Gesamtverhalten abgeben. Meistens werden nur zwei Antwortalternativen vorgegeben. Man muß z. B. angeben, ob eine Unternehmung in einem bestimmten Zeitraum insolvent wird oder nicht. Ein weiteres Merkmal dieser Studien ist, daß die zu beurteilende Unternehmensentwicklung ein historisch bereits stattgefundenes Ereignis darstellt, das von den Versuchspersonen nicht mehr beeinflußt werden kann.

Die Angemessenheit eines Gesamturteils spielt eine zentrale Rolle in den empirischen Studien, die dem sogenannten "Linsen"-Modell (lens-model) des Psychologen Brunswick folgen. Dieser Ansatz ist in der psychologischen Entscheidungsforschung weit verbreitet.³⁶⁾ Bei unterschiedlichen Beurteilungsproblemen konnte nachgewiesen werden,

- * daß statistische Modelle genauer klassifizieren als menschliche Beurteiler ("clinical vs. statistical prediction"),³⁷⁾

- * daß lineare Modelle, die das menschliche Urteilsverhalten "paramorph" abbilden, genauer klassifizieren als die modellierten Urteiler selbst ("man vs. model of man"),³⁸⁾ und

- * daß Mehrheitsvoten statistisch gebildeter Gruppen von Beurteilern besser klassifizieren als einzelne Beurteiler ("composite judge").³⁹⁾

Aufgrund dieser Forschungsergebnisse und betriebswirtschaftlicher Studien zur Insolvenzprognose auf Basis von Bilanzkennzahlen wurde mehrfach angeregt, den Menschen durch ein genaueres statistisches Klassifikationsmodell zu ersetzen. Libby vermutet insbesondere, daß Experten bei der Suche und Aufbereitung von Informationen überlegen seien, und schlägt daher eine Arbeitsteilung vor, bei der der Mensch die Informationsgewinnung und die Maschine die Informationsverarbeitung übernimmt.⁴⁰⁾

Libbys Vorschlag ist vor dem Hintergrund einer ganzen Reihe von Nachfolgestudien der englischsprachigen verhaltensorientierten Rechnungswesenforschung zu sehen, die das Linsen-Modell als Forschungsparadigma übernommen haben.⁴¹⁾ Von besonderer Bedeutung für unsere Arbeit sind jene Studien, in denen die Abgabe von Kreditwürdigkeitsurteilen erforscht wird.

In diesen Arbeiten werden den Kreditmanagern für ein Sample von ca. 30 bis 50 erfolgreichen und erfolglosen Unternehmen üblicherweise je drei bis sieben Kennzahlen vorgelegt, mit deren Hilfe sie die zukünftige Entwicklung der Fälle einstufen sollen. Die in Abbildung 3.2. zusammengefaßten Befunde belegen, daß statistische Modelle genauere Prognosen liefern als der durchschnittliche Analytiker. Man erkennt jedoch auch, daß die Überlegenheit deutlich abnimmt, wenn die Klassifikationsgüte des statistischen Modells an einem anderen Sample oder mit Hilfe eines Hold-Out-Verfahrens validiert wird. Bei praktischen Anwendungen ist außerdem zu berücksichtigen, daß die Diskriminanzfunktionen zeitliche Instabilitäten aufweisen. Da man bei einer ex-ante-Prognose nicht weiß, im wievielten Jahr vor einer eventuellen Insolvenz man sich befindet, kann die Klassifikationsgüte des Modells noch geringer als in Abbildung 3.2. ausfallen, weil man eine "unzeitgemäße" Prognosefunktion verwendet.

Die menschlichen Beurteiler schneiden **besser** ab als das validierte statistische Modell, wenn man nicht den einzelnen Analytiker, sondern ein statistisches Mehrheitsvotum entscheiden läßt. Bemerkenswert ist die deutliche Überlegenheit einer **natürlichen** Gruppe in der Untersuchung von Chalos. Nach seinen Aussagen ist ein derartiges "review committee" ein Standardverfahren in amerikanischen Banken. Wegen der kleinen Zahl beurteilter Fälle sollte man seinen Befund allerdings nicht ohne weiteres verallgemeinern.

Abbildung 3.2.: Klassifikationsgüten menschlicher Analytiker und statistischer Modelle bei Insolvenzprognosen

Quelle	Menschliche Analytiker				Statistisches Modell	
	Reales Einzel-Urteil	Modell. Einzel-Urteil	Statische Gruppe	Natürliche Gruppe	ohne Validierung	mit Validierung
Libby (Prediction 1975) und Libby (Model of Man 1976)	74,0 v.H.	72,2 v.H.	81,7 v.H.	---	85,0 v.H.	71,7 v.H.
Abdel-khalik/El Sheshai (Information Choice 1980)	62,5 v.H.	62,5 v.H.	---	---	95,3 v.H.	78,0 v.H.
Casey (Usefulness 1980) vgl. Casey (Information Load 1980)	56,7 v.H.	---	60,0 v.H.	---	73,0 bis 80,0 v.H.	---
Zimmer (Lens Study 1980)	77,4 v.H.	---	85,7 v.H.	---	88,1 v.H.	83,3 v.H.
Casey (Probability Disclosure 1983)	82,0 v.H.	---	---	---	96,0 v.H.	---
Chalos (Financial Distress 1985)	79,0 v.H.	---	92,0 v.H.	98,0 v.H.	89,0 v.H.	86,1 v.H.

Anmerkung: Bezüglich Aufgabenstellung, Effizienzmaß und Informationsangebot der Studien vgl. Abbildung 3.7.

Die Bedenken gegenüber dem Vorschlag von Libby wachsen, wenn man die Ergebnisse von Studien berücksichtigt, die sich der Feldsituation stärker annähern, indem sie ein größeres Informationsangebot oder (einfache) Entscheidungshilfen zur Verfügung stellen:

- * In der Studie von Abdel-khalik und El-Sheshai können die Kreditmanager zunächst **unterschiedliche** Mengen von Informationen beschaffen, um sie dann zu einem Urteil zu verknüpfen. Es zeigt sich, daß eine verbesserte, maschinelle Informationsverknüpfung auf Basis der beschafften Informationen lediglich zu einer Effizienzsteigerung um 5% führt. Mit einem Wechsel der Beschaffungsstrategie kann jedoch eine Steigerung um 23% erzielt werden.⁴²⁾
- * Ashton kann in einer Feldstudie mit unbegrenzten Nachfragemöglichkeiten nachweisen, daß die Manager des time Magazins für die vierteljährlich nachgefragten Anzeigenseiten **bessere** Prognosen als ein Regressionsmodell liefern, wenn man ihre Prognosen um einen systematischen Unterschätzungs-Bias bereinigt, der darauf beruht, daß die Konsequenzen zu optimistischer Prognosen gravierender bewertet werden als diejenigen zu pessimistischer Prognosen.⁴³⁾
- * Casey gibt einer Gruppe von Kreditmanagern sechs ausgewählte Kennzahlen, die nach einem statistischen Modell prognosetauglich sind, und zwar für das fünfte bis dritte Jahr vor Insolvenz. Eine zweite Gruppe erhält zusätzlich Bilanz- und Erfolgsrechnungen, eine dritte zusätzlich Erläuterungen des Bilanzierenden ("footnotes"). Die Gruppen mit Zusatzinformationen erreichen signifikant bessere Klassifikationsleistungen.⁴⁴⁾ Der Befund weist darauf hin, daß menschliche Analytiker Bilanzinformationen anders nutzen als statistische Modelle, denn in dem verwendeten Modell hatten die Zusatzinformationen keine signifikante Leistungssteigerung bewirkt.
- * Ebert und Kruse lassen ihr Informationsangebot von Experten der Wertpapieranalyse zusammenstellen. Es ergibt sich eine nach Meinung der Praktiker "unverzichtbare Mindestmenge" von 22 Informationen. Die Autoren

stellen fest, daß das Durchschnittsurteil der untersuchten fünf Analytiker ("composite judge") genauer ist als die Prognose des statistischen Modells auf der gleichen Datenbasis.⁴⁵⁾

* Mit einer sehr **einfachen graphischen Entscheidungshilfe**, den sog. "Schematischen Gesichtern", erreichen Studenten und Kreditmanager eine praktisch genauso hohe Klassifikationsgüte wie ein diskriminanzanalytisches Modell.⁴⁶⁾

Diese Studien weisen darauf hin, daß man auch Ergebnisse von Laborstudien aus anderen Forschungsdisziplinen, die häufig bestätigt wurden, nicht ohne weiteres auf betriebswirtschaftliche Feldsituationen übertragen sollte, weil **durchgängig** bestimmte Faktoren in der experimentellen Welt konstant gehalten wurden. Die Linsen-Modelle eignen sich möglicherweise dazu, Handlungsempfehlungen für eine **Entschlußphase** abzuleiten, in der nur noch wenige vorhandene Schlüsselinformationen optimal zu einem Gesamturteil verknüpft werden sollen. Sie können aber nicht erklären, wie man zu diesen Schlüsselinformationen kommt und mit den damit verbundenen Informationsproblemen fertig wird.⁴⁷⁾ Angesichts der großen Zahl von Linsen-Studien und der sehr geringen Zahl von prozessual angelegten Effizienzprüfungen, ist kritisch anzumerken, daß man zugunsten eines einfachen und klaren Effizienzmaßstabs "Genauigkeit des Gesamturteiles" und der Möglichkeit, alle Beziehungen im Rahmen einer geschlossenen Modellrechnung analysieren zu können, häufig das eigentliche Informationsproblem hinausdefiniert hat.

In diesem Zusammenhang ist auf eine wichtige Abgrenzung unserer Arbeit hinzuweisen: Linsen-Modell-Studien und Untersuchungen zur Bestimmung subjektiver Wahrscheinlichkeiten werden nicht nur in der psychologischen Entscheidungsforschung sondern auch in der verhaltensorientierten Rechnungswesenforschung häufig verwendet, um ein "beschränktes", "verzerrtes" Informationsverhalten des Menschen nachzuweisen.⁴⁸⁾ Derartige "Bias"-Studien sind Gegenstand heftiger Kontroversen.⁴⁹⁾ Wir weisen darauf hin, daß der "Bias"-Ansatz nicht mit dem von uns vorgeschlagenen Insuffizienzkonzept zu verwechseln ist.

Im "Bias"-Ansatz werden Entscheidungsträger als suboptimal dargestellt, weil sie Heuristiken einsetzen, die bei einfachen Entscheidungsproblemen sog. "normativen" Modellen unterlegen sind. Diese Heuristiken können aber bei komplexen Entscheidungen durchaus "rational" sein.⁵⁰⁾ Unser Insuffizienzkonzept richtet sich von vornherein auf komplexe Entscheidungen. Wir vergleichen nicht optimale mit suboptimalen Lösungen, sondern wollen Handlungsempfehlungen entwickeln, wie man schlechte Entscheidungen vermeiden kann. Wir wollen nicht nachweisen, daß Entscheidungsträger suboptimal handeln, sondern wir wollen **erklären, warum** es zu besonders schlechten Lösungen kommen kann, um aufgrund dieser **Ursachenanalyse** gezielte Handlungsempfehlungen abzuleiten. Deshalb ist bei uns das Informationsverhalten auch keine "black box", sondern der eigentliche Forschungsgegenstand.

Hier sehen wir das eigentliche Defizit des "man vs. model"-Vergleichs: Man stellt der Modellempfehlung nur den **durchschnittlichen** Analytiker gegenüber. Man vernachlässigt die großen interindividuellen **Unterschiede** und fragt nicht, wie sie zustande kommen. Damit nimmt man sich die Möglichkeit, systematisch von Experten zu lernen und Hinweise für eine gezielte Schulung ungeübter Benutzer zu erhalten. Die Entwicklung von Klassifikationsmodellen ist nur ein Weg um die Effizienz der Kreditentscheidung zu verbessern. Ein zweiter Weg besteht in der Erforschung des Informationsverhaltens, ein dritter, langfristig anzustrebender, liegt in der Verknüpfung beider Anstrengungen. Hierzu muß jedoch erst einmal der zweite Weg begangen werden.

Trotz der methodischen Kritik an Linsen-Modellen könnte man auch in unserem Fall vorsehen, den Nutzen der Bilanzanalyse an der **Qualität des Gesamturteils** zu messen. Wir bestreiten dies nicht grundsätzlich, wollen aber den Stellenwert eines derartigen Effizienzmaßes für unsere Aufgabenstellung erheblich relativieren:

1. Ein **"angemessenes"** Gesamturteil setzt voraus, daß es ein intersubjektiv **eindeutig richtiges Gesamturteil** gibt. Diese Voraussetzung ist selbst dann nicht zwingend gegeben, wenn man lediglich klassifizieren läßt, ob eine Firma innerhalb eines bestimmten Zeitraumes insolvent wurde oder nicht, weil die zu beurteilende Firma gerade wegen einer Kreditverweigerung insolvent geworden sein kann. In unseren experimentellen Fällen

ging es jedoch nicht um einen derartigen Schwarz-Weiß-Vergleich, sondern um eine Abstufung von Grautönen: Die zu untersuchenden Unternehmen sind in der Wirtschaftspresse bekanntgewordene Krisenfälle, bei denen unterschiedliche Arten von Krisenursachen vorgelegen haben.⁵¹⁾ Wie "gravierend" diese Krisen waren, läßt sich wohl intersubjektiv kaum eindeutig festlegen. Wir könnten demnach höchstens eine **Übereinstimmung mit einer Mehrheitsmeinung** von "Experten" feststellen.⁵²⁾

2. Wenn man sich wegen dieser intersubjektiven Divergenzen nur auf eine recht grobe Abstufung beschränkt, für die ein hinreichender Konsens vorliegt, dann wächst die Gefahr, daß ein Analytiker **zufällig die angemessene Kategorie** trifft. Da wir unserer Effizienzmessung wegen der Aufwendigkeit der Analysen nur vier Fälle zugrunde legen können, ist diese Gefahr relativ groß. Bei fünf Urteilstkategorien und vier mittelmäßig ausgeprägten Fällen beträgt die Wahrscheinlichkeit, daß man bei drei von vier zufällig eingestuften Fällen höchstens um eine Kategorie von der angemessenen abweicht, immerhin 0,6048.
3. Man kann nicht generell davon ausgehen, daß der Adressat der Bilanzanalyse, für den die Entscheidungsvorbereitung vorgenommen wird, und der von uns bewertete Analytiker die **Teilurteile in gleicher Weise gewichten**. Erstens können sich ihre Ziele und Risikopräferenzen unterscheiden, zweitens können sie über unterschiedliche Hintergrundinformationen verfügen. Der Adressat wird sich daher nicht auf das Gesamturteil verlassen wollen, sondern auch berücksichtigen, welche Teilurteile herangezogen werden und wie man diese ableitet und gewichtet. Deshalb sollte sich die Effizienzmessung (auch) auf die Qualität und Relevanz der abgegebenen Teilurteile und nicht (nur) auf das Gesamturteil stützen.⁵³⁾
4. Bilanzanalyse und Bilanzkritik dienen der **Überprüfung der Einhaltung vielfältiger Verhaltenserwartungen** verschiedener Bezugsgruppen.⁵⁴⁾ Wenn sich die Verhaltenserwartungen auf eine einzige Schlüsselkennzahl konzentrieren, hat der Bilanzersteller eine größere Möglichkeit, diese in seinem Sinne bilanzpolitisch zu beeinflussen. Formuliert der Bilanzadressat aber ein ganzes Netzwerk von Rollenerwartungen, die sich gegenseitig kontrollieren und ergänzen, dann wird eine vom Adressaten uner-

wünschte Beeinflussung erheblich erschwert. Für die Effizienzmessung bedeutet dies, daß die Güte der Bilanzanalyse auch an der adäquaten Überprüfung eines solchen Netzwerkes von Verhaltenserwartungen gemessen werden sollte.

5. Unterschiedliche Unternehmenssituationen erfordern auch **unterschiedliche Maßnahmen** zur Kreditsicherung. Eine differenzierte Bilanzanalyse kann eher gezielte Hinweise geben, welche Maßnahmen zu ergreifen sind als eine undifferenzierte.

Aus diesen Gründen wollen wir das Gesamturteil **nicht** als ein Effizienzmaß verwenden. Es widerspricht nicht zuletzt unserer Auffassung, daß Bilanzanalysen bessere Entscheidungen **vorbereiten** und nicht vorschnelle Entschlüsse rechtfertigen sollen. Mit Blick auf die Bilanzanalyse ist somit bereits die Forschungsfrage der Linsen-Modell-Studien falsch gestellt.

Abbildung 3.3.: Maßgrößen des Ressourcenverbrauches (Zeitaufwand) in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Hofstede (Behavioral Parameters 1972), S. 685	simulierte Anlageentscheidung, Gruppenmittelwerte zwischen 24 und 45 Minuten	Geschäftsbericht, Erfolgsrechnungen zweier Jahre, Prüferbericht, Fußnoten zum "financial statement", zusammen 5 Seiten
Kozar (Decision Making 1972), S. 82	Unternehmensspiel, Entscheidungszeiten pro Spielperiode, Gruppenmittelwerte reichen von 5 Minuten am Anfang bis unter 1 Minute am Ende	Berichte zur Produktionssteuerung: Materiallager, Fertigwarenlager, Kapazitäten, Kosten etc.
Mock/Estrin/Vasarhelyi (Learning 1972), S. 133, 145, S. 149	Unternehmensspiel, Entscheidungszeiten pro Spielperiode, Gruppenmittelwerte liegen zwischen 11 Minuten am Anfang und 3 Minuten am Ende, außerdem die Veränderungen der Entscheidungszeiten von Periode zu Periode	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: direkte vs. verzögerte Informationen über die erzielten Preise
Mock (Budget Information 1973), S. 527	Unternehmensspiel, Entscheidungszeiten pro Spielperiode, Gruppenmittelwerte liegen zwischen 45 Minuten am Anfang und 6 Minuten am Ende, außerdem die Veränderungen der Entscheidungszeiten von Periode zu Periode	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: Erwartungen und berichtete Budgetabweichung vs. keine Erwartung u. Abweichung
Chervany/Dickson (Overload 1974), S. 1339, 1340, 1342	Entscheidungszeit in einem Unternehmensspiel: Gruppenmittelwerte liegen zwischen 18,3 und 23,0 Minuten pro Spielperiode	exp. Bed.: verdichtete vs. unverdichtete Berichte (jeweils fünf) mit 449 bzw. 1040 Angebotsinformationen
Senn/Dickson (Purchasing 1974), S. 61	simulierte Beschaffungsentscheidung, mittlere Entscheidungszeit 15 Minuten, Abweichungen zwischen den Gruppen bis zu 10 Minuten	Informationen über Bedarf, Bestand, Kosten und Lieferanten, detailliert vs. aggregiert und hardcopy vs. Bildschirm
Driver/Mock (Information Processing 1975), S. 505	Unternehmensspiel, Entscheidungszeit pro Spielperiode: Gruppenmittelwerte liegen zwischen 52 und 67 Minuten in der ersten und zwischen 11 und 18 Minuten in der fünften Spielperiode	Daten über Produktion, Absatz und Finanzierung der letzten 10 Spielperioden, laufender Feedback über erzielten Erfolg; exp. Bed.: unterschiedlich detaillierte Kosten- und Leistungsberichte
Smith (Inquiry Techniques 1975), S. 276	Unternehmensspiel, Entscheidungszeiten pro Spielperiode: Gruppenmittelwerte liegen zwischen 132 und 163 Minuten in der ersten und zwischen 25 und 30 Minuten in der sechsten Spielperiode	exp. Bed.: Abfrage von bis zu 17 Berichten vs. zusätzliche individ. aufgebauete Berichte a) tabellarisch, b) in (grober) graphischer Form
Benbasat/Schroeder (MIS Design 1977), S. 41, 43	Unternehmensspiel, Entscheidungszeit, 60 % höherer Zeitverbrauch bei Einsatz von Entscheidungshilfen, ohne absolute Zeitangabe	exp. Bed.: 8 vs. 19 Berichte, außerdem tabellarische vs. graphische Darstellung, drei verschiedene Entscheidungshilfen
Jenkins (Design Variables 1977), S. 89, S. 174	Entscheidungszeit in einem Unternehmensspiel: Gruppenmittelwerte liegen zwischen 35 und 47 Minuten pro Spielperiode	exp. Bed.: Art der Zielvorgabe, Angebotsmedium: Bildschirm vs. EDV-Liste, verschiedene Berichte nachfragbar
Benbasat/Dexter (Value and Events 1979), S. 740, 746	Entscheidungszeit in einem Unternehmensspiel: Gruppenmittelwerte zwischen 56 und 83 Minuten	exp. Bed.: fertige Berichte vs. einzelne Informationen aus einer Datenbank
Brockhoff (Delphi-Prognosen 1979), S. 38	Prognose von Zinssätzen auf kurze Sicht (drei bzw. sechs Monate), Bearbeitungszeiten für 4 bis 7 Prognosefragen, Werte der einzelnen Delphi-Gruppen liegen zwischen 151 und 218 Minuten	87 verschiedene Zeitreihen, die in unterschiedlicher Präsentationsform aus einer Datenbank abgerufen werden können, außerdem werden die Schätzungen anderer Experten im Rahmen einer via Computerkonferenz realisierten Delphi-Gruppe mitgeteilt

Abbildung 3.3.: Maßgrößen des Ressourcenverbrauches (Zeitaufwand) in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Moriarity (Graphics 1979), S. 211 f., 215, 217	Entscheidungszeiten für die Beurteilung, ob ein Unternehmen insolvent wird oder nicht, 22 Unternehmen, Gruppenmittelwerte liegen zwischen 3 und 8 Minuten, Honorar der Versuchspersonen hängt von Trefferzahl und Zeitaufwand ab	vier exp. Bed.: schematische Gesichter a) ohne, b) mit Erklärung, c) 10 ausgewählte Daten aus Bilanz und Erfolgsrechnung, d) 13 ausgewählte Bilanzkennzahlen; alle Daten für sechs Jahre vor dem Insolvenzjahr
Casey (Information Load 1980), S. 40, 44	Beurteilung, ob ein Unternehmen in den nächsten drei Jahren insolvent wird, Gruppenmittelwerte zwischen 149 und 352 Minuten	3 exp. Bed.: a) sechs ausgewählte Bilanzkennzahlen, b) zusätzlich Bilanz und Erfolgsrechnung, c) zusätzlich Erläuterungen
Davis (Information Presentation 1980) S. 52, 55	Entscheidungszeit in einem Unternehmensspiel: Gruppenmittelwerte liegen zwischen 4,9 und 6,7 Minuten pro Spielperiode	exp. Bed.: 1. verdichtete vs. unverdichtete Berichte, 2. tabellarische vs. graphische Darstellung
White (Aggregation 1981), S. 63 f., S. 95f.	Entscheidungszeit bei einem Beurteilungsproblem: Gruppenmittelwerte liegen zwischen 12 und 44 Minuten	Feldexp.: Beurteilung unterschiedlich stark aggregierter Revisions-Berichte
Benbasat/Dexter (Individual Differences 1982), S. 7	Entscheidungszeit in einem Unternehmensspiel: Gruppenmittelwerte zwischen 48 und 80 Minuten	exp. Bed.: fertige Berichte vs. einzelne Informationen aus einer Datenbank
Otley/Dias (Aggregation 1982), S. 177 f.	simulierte Produktprogrammplanung, Entscheidungszeiten pro Spielperiode, Gruppenmittelwerte zwischen 75 Minuten am Anfang und 9 Minuten am Ende	unterschiedlich ausführliche Berichte über Produktion, Absatz und Erfolg, vier exp. Aggregationsbedingungen
Biggs (Financial Analysts 1984), S. 316	simulierte Anlageentscheidung, individuelle Zeiten zwischen 26 und 282 Minuten	35 bis 37 Datenzeilen aus Bilanzen und Erfolgsrechnungen, 5 Unternehmen, 10 Jahre
Jacoby/Mazursky/Troutman/Kuß (Feedback 1984), S. 535	simulierter Aktienkauf, individuelle Zeiten für alle Perioden zwischen 90 und 570 Minuten	26 verschiedene Bilanzkennzahlen und Börseninformationen für 8 Handelsunternehmen, historische Daten
Mock/Vasarhelyi (Cognitive Style 1984), S. 68 ff., Experiment IS4	Unternehmensspiel mit den Aktionsparametern Werbung, Produktionsmenge, Materialeinkauf, Entscheidungszeiten pro Spielperiode, Gruppenmittelwerte liegen zwischen 12 und 39 Minuten	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: unterschiedl. feines Feedback, Vollständigkeit der berichteten Budgetabweichung
Benbasat/Dexter (Graphical Information Presentation 1985), S. 1360	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, Entscheidungszeiten für 10 Spielperioden, Gruppenmittelwerte liegen zwischen 27 und 37 Minuten	Marketingaufwand und erzielte Gewinne für drei Verkaufsgebiete; exp. Bed.: a) Tabelle vs. Graphik, b) mono- vs. multichromatisch; außerdem zufällige Zuteilung von Fällen und Messung kogn. Stile (GEFT)
Benbasat/Dexter (Time Constraints 1985), S. 34	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, Entscheidungszeiten für 30 Simulationen a) bei exp. Bed. max. 15 Min.: Gruppenmittelwerte liegen zwischen 10 und 15 Minuten, b) bei exp. Bed. max. 5 Min.: Gruppenmittelwerte liegen zwischen 4 und 6 Minuten (bis Aufgabenlösung)	Marketingaufwand und erzielte Gewinne für drei Verkaufsgebiete; exp. Bed.: a) Tabelle vs. Graphik vs. Graphik und Tabelle, b) mono- vs. multichromatisch, c) enge vs. weite Zeitbeschränkung (5 vs. 15 Min.),
Benbasat/Dexter/Todd (Graphical Information Presentation 1985), S. 35	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, Zeiten zum Studieren des Informationsangebotes, Gruppenmittelwerte (mittlere Zeiten) zwischen 77 und 135 Sekunden	Marketingaufwand und erzielte Gewinne für drei Verkaufsgebiete; exp. Bed.: a) Tabelle vs. Graphik, b) mono- vs. multichromatisch, c) Fälle mit niedrigem vs. hohem Marketingbudget

Abbildung 3.4.: Maßgrößen der subjektiv wahrgenommenen Leistung in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Bronner (Zeitdruck 1973), S. 127, 161	Unternehmensspiel, Leistungszufriedenheit: "Wie beurteilen Sie den Arbeitserfolg der heutigen Sitzung?", "Wie beurteilen Sie die Zusammenarbeit (a) innerhalb Ihrer eigenen Gruppe, (b) mit Ihrem Stab bzw. Vorstand?"	Die Spielgruppen bestehen aus den räumlich getrennten Teilen "Stab" und "Vorstand". Sie müssen schriftlich miteinander kommunizieren. Nur der Stab erhält die Routineberichte zum Betriebs- und Finanzbereich. Er kann zusätzliche gebührenpflichtige Informationen von der Spielleitung erfragen.
Chervany/Dickson (Overload 1974), S. 1339, 1340, 1342	Unternehmensspiel, Konfidenz in die getroffenen Produktions-, Personal- und Materialentscheidungen, siebenstufiges Rating	exp. Bed.: verdichtete vs. unverdichtete Berichte (jeweils fünf) mit 449 bzw. 1040 Angebotsinformationen
Dermer/Siegel (Behavioral Measures 1974) S. 92 f.	Unternehmensspiel, Zufriedenheit mit eigener Leistung, Gruppenleistung und Veranstaltung	Feedback über Spielleistung und Verhaltensbewertung (keine näheren Angaben)
Senn (Information System 1973), S. 103, S. 133 ff. vgl. auch Senn/Dickson (Purchasing 1974), S. 62	simulierte Beschaffungsentscheidung, subjektive Leistungseinschätzung, eigene Entscheidung wäre besser als x Prozent der Einkäufer	Informationen über Bedarf, Bestand, Kosten und Lieferanten, detailliert vs. aggregiert und hardcopy vs. Bildschirm
Börsig (Operations Research 1975), S. 119, 123	Unternehmensspiel, Adoption/Rejektion eines Operations-Research-Modells, bestimmt nach der Abweichung von Modellempfehlung und tatsächlicher Entscheidung, subjektive Bewertung von Entscheidungsmodellen	Firmen- und Rollenbeschreibung, betriebswirtschaftliche Informationen über Produktion, Absatz und Finanzierung
Börsig (Operations Research 1975), S. 124 f.	Unternehmensspiel, Zufriedenheit mit der eigenen Leistung und der Gruppenleistung, Bewertung des Entscheidungsverlaufes, Zufriedenheit mit der getroffenen Entscheidung	Firmen- und Rollenbeschreibung, betriebswirtschaftliche Informationen über Produktion, Absatz und Finanzierung
Libby (Decision Makers 1975), S. 481 vgl. auch Libby (Prediction 1975), S. 158	Konfidenz in die Klassifikation, ob ein Unternehmen innerhalb der nächsten drei Jahre insolvent wird oder nicht, dreistufige Ratingskala	fünf Bilanzkennzahlen für eines der drei Jahre vor der Insolvenz
Schroeder/Benbasat (Uncertainty 1975), S. 538	Unternehmensspiel, Beschaffung und Lagerhaltg., Konfidenz in die getroffene Entscheidung, siebenstufiges Rating	gebührenpflichtige Beschaffung von bis zu 19 Berichten mit versch. Umfang und Aggregationsgraden; exp. Bed.: Unsicherheit der Nachfrageentwicklung
Brockhoff (Delphi-Prognosen 1979), S. 85, 193 Vgl. auch Brockhoff (Experimente 1977), S. 513	Verschiedene Fragen zur Beurteilung der Datenbank und der Dialogführung (ausreichende Informationen, verständliche Texte, Aktualität der Daten, Schwierigkeit der Systemhandhabung, Wunsch über ein ähnliches System am Arbeitsplatz zu verfügen, Beurteilung der Teilnahme am Experiment u. a.)	87 verschiedene Zeitreihen, die in unterschiedlicher Präsentationsform aus einer Datenbank abgerufen werden können, außerdem werden die Schätzungen anderer Experten im Rahmen einer via Computerkonferenz realisierten Delphi-Gruppe mitgeteilt
Picot/Lange (Synoptische versus inkrementale Gestaltung 1979), S. 585f., 588f.	Unternehmensspiel, zahlreiche Einzelfragen zur Gruppenleistung und zur eigenen Rolle, "begleitend" nach den Spielperioden erfragt, weitere Fragen nach Abschluß des Spieles	Ausführliche Betriebs-, Finanz- und Vertriebsberichte; exp. Bed.: verdeckte Vorgabe von inkrementaler und synoptischer Planungsstrategie

Abbildung 3.4.: Maßgrößen der subjektiv wahrgenommenen Leistung in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Schütt (Wahrscheinlichkeitsschätzungen 1979), S. 218ff., 271 ff., 306 ff.	Zufriedenheit mit angebotenen Verfahren zur Bildung subjektiver Wahrscheinlichkeiten und mit eigener Leistung	vereinfachtes, fiktives Entscheidungsproblem, Verwendung verschiedener Informations- und Entscheidungshilfen
Casey (Usefulness 1980), S. 610	Konfidenz in die Beurteilung, ob das Unternehmen in den nächsten drei Jahren insolvent wird	sechs ausgewählte Kennzahlen, fünftes bis drittes Jahr vor der Insolvenz
Uebele (Planungstechniken 1980), S. 201 f.	simulierte Preisentscheidungen, Akzeptanz des Entscheidungsbaumes ("würde ... verwenden" vs. "nicht verwenden"), ausreichende Informationsgrundlage, Angemessenheit der Vorgehensweise	Feedback über Umweltreaktionen, Absatzmengen, Erfolg, ab der sechsten von zehn Spielperioden obligat. Verwendung eines standardisierten Entscheidungsbaumes
Uebele (Planungstechniken 1980), S. 234	simulierte Preisentscheidungen, Zufriedenheit mit dem Entscheidungsprozeß: "Wie sind Sie mit dem Ablauf der Preisentscheidungsprozesse ... zufrieden?"	Feedback über Umweltreaktionen, Absatzmengen, Erfolg, ab der sechsten von zehn Spielperioden obligat. Verwendung eines standardisierten Entscheidungsbaumes
Zimmer (Lens Study 1980), S. 631	Konfidenz in die Klassifikation, ob ein Unternehmen im nächsten Jahr insolvent wird	fünf Bilanzkennzahlen für drei Jahre vor der Insolvenz
Davis (Information Presentation 1981) S. 38	Konfidenz in die getroffene Entscheidung: 1: "no confidence" bis 10: "utmost confidence"	exp. Bed.: 1. verdichtete vs. unverdichtete Berichte, 2. tabellarische vs. graphische Darstellung
Lange (Portfolio-Methoden 1981), S. 236 f., S. 269	Fallstudie zur strategischen Planung, sehr stark vereinfacht, "Sind Sie insgesamt mit der Planungsmethode zufrieden" 1: "in schwachem Maße" bis 7: "in starkem Maße", zahlr. weitere Fragen zur Bewertung der Methode	ausgewählte Branchen- und Firmendaten verschiedener unabh. Märkte; exp. Bed.: drei versch. Planungsmethoden, eine Kontrollgr.
Lucas (Graphics 1981), S. 762	simulierte Bestellmengenentscheidung, Zufriedenheitsmaß, Mittelwert aus Konfidenz in die getroffene Entscheidung und Zufriedenheit mit der eigenen Leistung	Kostenangaben, Kapazitätsbeschränkungen, Verteilungen von Nachfrage-, Bestands- und Fehlmengen; exp. Bed.: tabellarische und graphische Präsentation
White (Aggregation 1981), S. 95 ff., S. 170	Konfidenz: "how confident are you in your solution?" 1: "not confident" 7: "very confident"	Feldexp.: Beurteilung unterschiedlich stark aggregierter Revisions-Berichte
Leichtfuß/Schmidt (Efficiency 1983), S. 10	Unternehmensspiel zur Abstimmung von Investitions- und Finanzierungsentscheidungen in einer divisionalisierten Unternehmung, Zufriedenheit mit dem Abstimmungsprozeß, Übereinstimmung mit der erreichten Lösung, Verständlichkeit des Algorithmus und des Informationsaustausches, wahrgenommene Entschlußautonomie und Partizipation an der Entscheidung	Firmen- und Rollenbeschreibung, Informationen über Investitionsprojekte, Finanzierungsmöglichkeiten und Budgetrestriktionen; exp. Bed.: Computerkonferenz vs. Empfehlung eines Algorithmus vs. reinalgorithmische Dekompositionslösung
Danos/Holt/Imhoff (Bond Raters 1984), S. 556	Konfidenz in das Bond-Rating, Entropie der subjektiven Wahrscheinlichkeitsverteilung über die möglichen Ratingwerte	Hintergrundinformationen über Unternehmen und Branche, 5 ausgewählte Bilanzkennzahlen für die letzten 5 Jahre, danach zusätzlich: Prognosen dieser Bilanzkennzahlen für die nächsten 5 Jahre

Abbildung 3.5.: Maßgrößen des ökonomischen Erfolges in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Mock (Information Structures 1969), S. 125	Unternehmensspiel mit den Aktionsparametern Werbung, Produktionsmenge, Materialeinkauf, erzielte Gewinne und Kosten, absolut und in Prozent der optimalen Lösung	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: direkte vs. verzögerte Preisinformationen
Dalton/Miner (Accounting 1970), S. 136 f.	Unternehmensspiel, 100-Punkte-Skala, basierend auf Gewinnen, Liquiditätsplanungs- und Lagerhaltungseffizienz	Gewinn- und Verlustrechnungen, Bilanzen, Produktions- und Absatzberichte, Informationen über Branche und Konkurrenten
Pankoff/Virgil (Usefulness 1970), S. 3, 19	Börsenspiel, erzielte Rendite nach Abzug der Informationsbeschaffungsgebühren, am Ende jeder Spielperiode wird liquidiert	35 Informationen über gesamtwirtschaftliche Lage, Branche und Unternehmen für je drei Jahre und drei Firmen, kostenpflichtige Beschaffung, historische Daten
Bronner/Witte/Wossidlo (Experimente 1972), S. 198	Unternehmensspiel, erzielter Gesamterfolg über 10 Spielperioden (Erträge - Aufwendungen), Index der vom Computermodell errechneten Produktqualität am Ende der letzten Spielperiode	Die Spielgruppen bestehen aus den räumlich getrennten Teilen "Stab" und "Vorstand". Sie müssen schriftlich miteinander kommunizieren. Nur der Stab erhält die Routineberichte zum Betriebs- und Finanzbereich. Er kann zusätzlich gebührenpflichtige Informationen von der Spielleitung erfragen.
Ebert (Decision Effectiveness 1972), S. 439, 441-444	Unternehmensspiel zur Abstimmung von Produktions- und Absatzmengen, Kosten für Personalwechsel, Überstunden/Fehlzeiten, Lagerhaltung und Fehlmengen, verschiedene Methoden zur Bestimmung der Effizienzwerte	Feedback über getroffene Entscheidungen, Kosten, Lagerhaltung, Umsätze; exp. Bed.: formale vs. informale Planung, mit/ohne irrelevante Informationen
Kozar (Decision Making 1972), S. 52 f., S. 75, S. 79 ff.	Unternehmensspiel, erzielte Durchschnittskosten, erzielte Gesamtkosten	Berichte zur Produktionssteuerung: Materiallager, Fertigwarenlager, Kapazitäten, Kosten etc.
Mock/Estrin/Vasarhelyi (Learning 1972), S. 135, vgl. Mock (Information Structures 1969), S. 125	Unternehmensspiel, erzielte Gewinne und Kosten, absolut und in Prozent der optimalen Lösung, außerdem Veränderungen der prozentualen Effizienzwerte von Periode zu Periode	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: direkte vs. verzögerte Preisinformationen
Bollom (Interim Reporting 1973) S. 14, 18 f.	einfaches Börsenspiel, erzielter Return on Investment, starke Vereinfachungen hinsichtlich Transaktionskosten u. Alternativanlagen	Ausführliche Quartalsberichte; exp. Bed.: sechs versch. Berichtsvarianten, gekreuzt mit zwei konjunkturellen Verläufen
Bronner (Zeitdruck 1973), S. 127, 161	Unternehmensspiel, erzielter Gesamterfolg über 10 Spielperioden (Erträge - Aufwendungen), Rangplätze gegenüber den drei konkurrierenden Unternehmen bei fünf Einzelkriterien und 10 Spielperioden	Die Spielgruppen bestehen aus den räumlich getrennten Teilen "Stab" und "Vorstand". Sie müssen schriftlich miteinander kommunizieren. Nur der Stab erhält die Routineberichte zum Betriebs- und Finanzbereich. Er kann zusätzlich gebührenpflichtige Informationen von der Spielleitung erfragen.
Mock (Budget Information 1973), S. 522 f. vgl. Mock (Information Structures 1969) S. 125	Unternehmensspiel, erzielte Gewinne und Kosten, absolut und in Prozent der optimalen Lösung, außerdem Veränderungen der prozentualen Effizienzwerte von Periode zu Periode	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: Erwartungen und berichtete Budgetabweichung vs. keine Erwartung u. Abweichung
Chervany/Dickson (Overload 1974), S. 1337, 1340 f.	Unternehmensspiel, Summe der Kosten für Lohn, Personalwechsel, Material, Bestellungen, Lagerhaltung, Transport und Fehlmengen pro Periode	exp. Bed.: verdichtete vs. unverdichtete Berichte (jeweils fünf) mit 449 bzw. 1040 Angebotsinformationen

Abbildung 3.5.: Maßgrößen des ökonomischen Erfolges in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Dermer/Siegel (Behavioral Measures 1974) S. 92 f.	Unternehmensspiel, Aktienkurse, Umsätze, Gewinne	Feedback über erzielte Umsätze, Gewinne und Aktienkurse (keine näheren Angaben)
Senn (Information System 1973), S. 49 f., 102, 118 ff. vgl. auch Senn/Dickson (Purchasing 1974) S. 60	simulierte Beschaffungsentscheidung, Kosten für Lagerhaltung, Transport und Fehlmengen	Informationen über Bedarf, Bestand, Kosten und Lieferanten, detailliert vs. aggregiert und hardcopy vs. Bildschirm
Hamner/Carter (Decision Rules 1975), S. 328 ff.	Unternehmensspiel zur Abstimmung von Produktions- und Absatzmengen, Kosten für Personalwechsel, Überstunden/Fehlzeiten, Lagerhaltung und Fehlmengen, verschiedene Methoden zur Bestimmung der Effizienzwerte	Feedback über getroffene Entscheidungen, Kosten, Lagerhaltung, Umsätze; exp. Bed.: Zuverlässigkeit und Zeithorizont für die Prognose der Nachfrage
Moskowitz/Miller (Production Planning 1975) S. 361 ff.	Unternehmensspiel zur Abstimmung von Produktions- und Absatzmengen, Kosten für Personalwechsel, Überstunden/Fehlzeiten, Lagerhaltung und Fehlmengen, verschiedene Methoden zur Bestimmung der Effizienzwerte	Feedback über getroffene Entscheidungen, Kosten, Lagerhaltung, Umsätze; exp. Bed.: Zuverlässigkeit und Zeithorizont für die Prognose der Nachfrage
Schroeder/Benbasat (Uncertainty 1975), S. 538	Unternehmensspiel Beschaffung und Lagerhaltung, Gesamtkosten für Auftragsabwicklung, Nachlieferung, Fehlmengen, Transport und Informationsbeschaffung	gebührenpflichtige Beschaffung von bis zu 19 Berichten mit versch. Umfang und Aggregationsgraden; exp. Bed.: Unsicherheit der Nachfrageentwicklung
Smith (Inquiry Techniques 1975), S. 49 ff.	Unternehmensspiel, Summe der Kosten für Lohn, Personalwechsel, Material, Bestellungen, Lagerhaltung, Transport und Fehlmengen pro Periode	exp. Bed.: Abfrage von bis zu 17 Berichten vs. zusätzliche individ. aufgebaute Berichte: a) tabellarisch, b) in (grober) graphischer Form
Wynne/Dickson (Performance 1975), S. 193	simulierte Beschaffungsentscheidungen, kumulierte Kostenersparnisse gegenüber einer rein maschinellen Serie von Entscheidungen: a) unter Ausschluß von Risiken, b) basierend auf Strategieparametern der Versuchspersonen	Marktforschungs-, Kapazitätsinformationen, Feedback über Erfolg, Informationen über außergewöhnliche Einkaufsmöglichkeiten, weitere, nicht näher spezifizierte Informationsmöglichkeiten im Computer-Dialog
Driver/Mock (Experimental Results 1976) S. 16 ff.	Unternehmensspiel, erzielte Gewinne und Kosten, absolut und in Prozent der optimalen Lösung, außerdem Veränderungen der prozentualen Effizienzwerte von Periode zu Periode	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: Vollständigkeit des Feedbacks über Budgetabweichungen
San Miguel (Information Processing 1976), S. 363 f.	simulierte Entscheidung über Stilllegung oder Weiterführung eines Werkes, Abweichung des Erwartungswertes der gewählten von der optimalen Alternative, wobei versch. Ziele nach subj. Gewichten der Vpn aggregiert werden	gegen Gebühr können bis zu 15 Kennzahlen zur Erfolgs-, Finanz- und Vermögenslage und den sozialen Beziehungen nachgefragt werden; exp. Bed.: 5 versch. Grade der Umweltunsicherheit
Benbasat/Schroeder (MIS Design 1977), S.41	Unternehmensspiel, Summe aus Bestell-, Transport-, Lagerhaltungs- und Fehlmengenkosten	exp. Bed.: 8 vs. 19 Berichte, außerdem tabellarische vs. graphische Darstellung, drei verschiedene Entscheidungshilfen
Dentler (Problemlöseverhalten 1977) S. 137 ff. Tabelle 48 bis 54	Managementspiel, simuliertes Verhalten eines afrikanischen Entwicklungsgebietes; die Effizienzbeurteilung stützt sich auf die Entwicklung von Bevölkerung, Tierbeständen, landwirtschaftlichen Erträgen und Zeitverbrauch von Entwicklungsprojekten	Computerprogramm mit zahlreichen Variablen und Verknüpfungen, Versuchspersonen können deren Zustände von einem Versuchsleiter erfragen,

Abbildung 3.5.: Maßgrößen des ökonomischen Erfolges in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Jenkins (Design Variables 1977), S. 88	Unternehmensspiel aus dem Produktionsbereich, gesamte Produktionskosten ("total operating costs"), Kostenfunktionen werden erläutert	exp. Bed.: Art der Zielvorgabe, Angebotsmedium: Bildschirm vs. EDV-Liste, verschiedene Berichte nachfragbar
Dörner/Reither (Problemlösen 1978), S. 535 ff.	Managementspiel, simuliertes Verhalten eines afrikanischen Entwicklungsgebietes, die Effizienzbeurteilung stützt sich auf die Entwicklung von Bevölkerung, Tierbeständen und landwirtschaftlichen Erträgen	Computerprogramm mit zahlreichen Variablen und Verknüpfungen, Versuchspersonen können deren Zustände von einem Versuchsleiter erfragen, Antwort entspricht dem qualitativen Niveau der Frage
Remus (Competitive Gaming 1978), S. 829, 832	Unternehmensspiel, Oligopol-situation, Entscheidungen über Marketing, Produktion, Forschung und Entwicklungszeiten, Finanzierung, erzielter Return on Investment, Rangplatz gegenüber konkurrierenden Mitspielern	Informationen über Produktion, Marketing, Finanzierung, Markt, Branche, Wettbewerber und gesamtwirtschaftliche Lage
Benbasat/Dexter (Value and Events 1979), S. 740	Deckungsbeiträge in einem Unternehmensspiel mit Abstimmung von Produktions- und Absatzmengen	exp. Bed.: Nachfrage von fertigen Berichten oder einzelnen Informationen aus einer Datenbank
Picot/Lange (Synoptische versus inkrementale Gestaltung 1979), S. 584, 590, ausführlicher Picot/Lange (Wirkungsanalyse 1978), S. 40-43	Unternehmensspiel, erzielter Gesamterfolg über 13 Spielperioden, Return on Investment, Eigenkapitalrendite, Periodengewinne, Kapitalumschlag, Umsatzrentabilität	Ausführliche Betriebs-, Finanz- und Vertriebsberichte; exp. Bed.: verdeckte Vorgabe von inkrementaler und synoptischer Planungsstrategie
Remus/Carter/Jenicke (Decision Rules 1979), S. 192 ff.	Unternehmensspiel zur Abstimmung von Produktions- und Absatzmengen, Kosten für Personalwechsel, Überstunden/Fehlzeiten, Lagerhaltung und Fehlmengen, verschiedene Methoden zur Bestimmung der Effizienzwerte, außerdem Test auf Lernfortschritte	Feedback über getroffene Entscheidungen, Kosten, Lagerhaltung, Umsätze; exp. Bed.: ab der dreizehnten Periode andere Kostenstruktur, hierüber keine vs. grobe vs. präzise Information
Chorba/New (Information Support 1980), S. 609 f.	Unternehmensspiel, stark vereinfachtes Marketingspiel mit vier Wettbewerbern, kumulierte Einzahlungsüberschüsse	exp. Bed.: teilweise vs. vollständige Informationslieferung vs. auf Nachfrage gebührenpflichtige Informationslieferung maximal 92 lieferbare Items
Lucas/Nielsen (Presentation 1980), S. 984, 986 f.	simulierte Logistikkentscheidungen, versch. Transportmittel und Läger, erzielte Gewinne und deren Veränderung	3 Berichte über Transportstatistik, Lagerhaltung, Kosten und Leistungen; exp. Bed. zu Angebotsumfang, Präsentationsform und Ausgabemedium
Uebele (Planungstechniken 1980), S. 193	simulierte Preisentscheidungen, gewichteter kumulierter Gesamterfolg, modifiziert um Abschläge für Abweichungen von einem Soll-Erfolg, das Honorar der Versuchspersonen hängt vom Gesamterfolg ab	Feedback über Umweltreaktionen, Absatzmengen, Erfolg, ab der sechsten von zehn Spielperioden obligatorische Verwendung eines standardisierten Entscheidungsbau- mes, mit Erläuterungen eines fiktiven Stabes
Benbasat/Dexter/Masulis (Human/Computer Interface 1981), S. 756	Bestimmung einer gewinnmaximierenden Preis-Mengen-Kombination: a) ob optimale Kombination gefunden wird oder nicht, b) erreichter Gewinn	Information über eingegebene Preise und Mengen: sortierte, unsortierte Liste, graphische Darstellung; exp. Bed.: system- vs. benutzergesteuerte Dialogführung, Variationen der Kommandosprache
Lange (Portfolio-Methoden 1981), S. 234 f., S. 254	Fallstudie zur strategischen Planung, sehr stark vereinfacht, Return on Investment, Cash Flow	ausgewählte Branchen- und Firmendaten verschiedener unabh. Märkte; exp. Bed.: drei versch. Planungsmethoden, eine Kontrollgr.

Abbildung 3.5.: Maßgrößen des ökonomischen Erfolges in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Lucas (Graphics 1981), S. 760 f.	simulierte Bestellmengenentscheidung, Summe aus Lagerhaltungs- und Fehlmengenkosten, bestes Ergebnis aus acht, durchschnittliches Ergebnis aus zehn weiteren Simulationsläufen	Kostenangaben, Kapazitätsbeschränkungen, Wahrscheinlichkeitsverteilungen der saisonal schwankenden Nachfrage und resultierender Bestands- und Fehlmengen; exp. Bed.: tabellarische und graphische Präsentation
Davis (Information Presentation 1980) S. 35f.	Unternehmensspiel, Summe aus Produktionswechselkosten, Lagerhaltungskosten für Normal- u. Zusatzbestände, Fehlmengenkosten	exp. Bed.: 1. verdichtete vs. unverdichtete Berichte, 2. tabellarische vs. graphische Darstellung
Putz-Osterloh (Problemlöseerfolg 1981), S. 87	Unternehmensspiel, kleiner Textilhersteller, Anzahl der Monate mit steigendem Zahlungsmittelbestand (als "Kapital" bezeichnet)	System besteht aus 24 Variablen, deren Werte vom Versuchsleiter erfragt werden können, 11 davon sind Aktionsparameter; exp. Bed.: Art der Zielvorgabe und Hinweise auf Beziehungen zwischen den Variablen, sowie deren Startwerte
Putz-Osterloh/Lüer (Problemlöseleistungen 1981), S. 318	Unternehmensspiel, kleiner Textilhersteller, Anzahl der Monate mit steigendem Zahlungsmittelbestand (als "Betriebskapital" bezeichnet)	System besteht aus 24 Variablen, deren Werte vom Versuchsleiter erfragt werden können, 11 davon sind Aktionsparameter; exp. Bed.: nur verbale Instruktion vs. schaubildliche Darstellung der Variablen und ihrer Zusammenhänge
Benbasat/Dexter (Individual Differences 1982), S. 7	Deckungsbeiträge in einem Unternehmensspiel mit Abstimmung von Produktions- und Absatzmengen	exp. Bed.: ohne vs. mit Entscheidungshilfe (Simulation der eigenen Entscheidungen), vier summarische Berichte
Ghani/Lusk (Information Representation 1982), S. 273	Einfaches Unternehmensspiel, Absatzmengenplanung unter Unsicherheit, erzielte Gewinne	Feedback über Umsatz, Kosten, Lagerbestände; exp. Bed.: Tabelle vs. Graphik
McIntyre (Marketing Models 1982), S. 24	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, erzielte Gewinne, relativiert am Optimum; Vermeidung von Fehlentscheidungen, Anstieg und Gleichmäßigkeit der Gewinne im Zeitablauf	Vpn müssen subj. Reaktionsfunktionen schätzen und Budgetaufteilung auf versch. Verkaufsgebiete vornehmen, (vorher gleich) Feedback über erzielte Gewinne und Umsätze der drei bzw. sechs Verkaufsgebiete
Dörner/Kreuzig/Reither/Stäudel (Lohhausen 1983), S. 115 f., 143 ff., insbes. 162 f.	Managementspiel, simuliertes Verhalten einer mittelgroßen hessischen Gemeinde, die Beurteilung der Leistungsgüte stützt sich auf die Entwicklung von 17 Kernvariablen zur wirtschaftlichen und sozialen Entwicklung, Effizienzbeurteilung wird vielfältig validiert	Computerprogramm enthält rund 2.000 Variablen mit zahlreichen Verknüpfungen, Versuchspersonen können deren Zustände von einem Versuchsleiter erfragen, Antwort entspricht dem qualitativen Niveau der Frage
Funke (Testintelligenz 1983), S. 287, 294	Unternehmensspiel, kleiner Textilhersteller, Anzahl der Monate mit steigendem "Gesamt"-Kapital bzw. steigendem "Flüssig"-Kapital	Pro Spielperiode werden 20 Zustandsvariablen angeboten und Entscheidungen zu 11 Aktionsparametern erbeten
Leichtfuß/Schmidt (Efficiency 1983), S. 10	Unternehmensspiel zur Abstimmung von Investitions- und Finanzierungsentscheidungen in einer divisionalisierten Unternehmung, Endwert der Investitionen	Firmen- und Rollenbeschreibung, Informationen über Investitionsprojekte, Finanzierungsmöglichkeiten; exp. Bed.: Computerkonferenz vs. Empfehlung eines Algorithmus vs. rein algorithmische Lösung
Putz-Osterloh (Problemlöseleistung 1983), S. 107	Unternehmensspiel, kleiner Textilhersteller, Anzahl von Monaten, in denen die Problemlöser "Gewinn" erzielt haben	System besteht aus 24 Variablen, deren Werte vom Versuchsleiter erfragt werden können, 11 davon sind Aktionsparameter, exp. Bed.: positive Rückmeldung, Aufforderung zur Reflexion, Kontrollgruppe

Abbildung 3.5.: Maßgrößen des ökonomischen Erfolges in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Craft/Mock/Rowe/Ryan ("Decisive" Style 1984), S. 12 ff.	Unternehmensspiel mit den Aktionsparametern Werbung, Produktionsmenge, Materialeinkauf, erzielte Gewinne	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: a) Art des Feedbacks, b) Berichtskomplexität
Jacoby/Mazursky/Troutman/Kuß (Feedback 1984), S. 536	simulierter Aktienkauf, Entwicklung des um Dividenden bereinigten Aktienkurses, der beste Finanzanalyst erhält 500 Dollar	26 verschiedene Bilanzkennzahlen und Börseninformationen für 8 Handelsunternehmen, historische Daten, max. 104 nachfragbar
Mock/Vasarhelyi (Cognitive Style 1984), S. 68 ff., Experiment IS4	Unternehmensspiel mit den Aktionsparametern Werbung, Produktionsmenge, Materialeinkauf, erzielte Gewinne in Dollars	Firmen- und Rollenbeschreibung, Informationen über Produktion, Absatz, Kosten, Leistungen der letzten 10 Jahre; exp. Bed.: unterschiedl. feines Feedback, Vollständigkeit der berichteten Budgetabweichung
Mock/Vasarhelyi (Cognitive Style 1984), S. 68 ff., Experiment SX1	Börsenspiel, erzielte Renditen durch Kursveränderungen, Dividenden und Zinsen	laufende Kursnotierungen der New Yorker Börse von 50 Wertpapieren
Remus (Presentations 1984), S. 534, 538 f.	Unternehmensspiel zur Abstimmung von Produktions- und Absatzmengen, Kosten für Personalwechsel, Überstunden/Fehlzeiten, Lagerhaltung und Fehlmengen, verschiedene Methoden zur Bestimmung der Effizienzwerte	Umsatzprognosen für die nächsten drei Perioden, Lagermengen, Beschäftigte, Arbeitsproduktivität und Kosten; exp. Bed.: graphische vs. tabellarische Präsentation der Kosteninformationen
Benbasat/Dexter (Graphical Information Presentation 1985), S. 1352	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, erzielte Gewinne, relativiert am Optimum, vergleichsweise einfaches Optimierungsproblem	Marketingaufwand und erzielte Gewinne für drei Verkaufsgebiete; exp. Bed.: a) Tabelle vs. Graphik, b) mono- vs. multichromatisch; außerdem zufällige Zuteilung von Fällen und Messung kogn. Stile (GEFT)
Benbasat/Dexter (Time Constraints 1985), S. 9	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, erzielte Gewinne, relativiert am Optimum, vergleichsweise einfaches Optimierungsproblem	Marketingaufwand und erzielte Gewinne für drei Verkaufsgebiete; exp. Bed.: a) Tabelle vs. Graphik vs. Graphik und Tabelle, b) mono- vs. multichromatisch c) enge vs. weite Zeitbeschränkung (5 vs. 15 Min.),
Benbasat/Dexter/Todd (Graphical Information Presentation 1985), S. 7	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Modells von Lodish, erzielte Gewinne, relativiert am Optimum, vergleichsweise einfaches Optimierungsproblem	Marketingaufwand und erzielte Gewinne für drei Verkaufsgebiete; exp. Bed.: a) Tabelle vs. Graphik, b) mono- vs. multichromatisch, c) Fälle mit niedrigem vs. hohem Marketingbudget
Jacoby/Kuß/Mazursky/Troutman (Effectiveness 1985), S. 102	simulierter Aktienkauf, Entwicklung des um Dividenden bereinigten Aktienkurses, der beste Finanzanalyst erhält 500 Dollar	26 verschiedene Bilanzkennzahlen und Börseninformationen für 8 Handelsunternehmen, historische Daten, max. 104 nachfragbar
Putz-Osterloh (Selbstreflexionen 1985), S. 15, Tabelle 1 bis 5	Managementspiel, simuliertes Verhalten eines afrikanischen Entwicklungsgebietes, die Effizienzbeurteilung stützt sich auf die Entwicklung von Bevölkerung, Tierbeständen und landwirtschaftlichen Erträgen	Computerprogramm mit zahlreichen Variablen und Verknüpfungen, Versuchspersonen können deren Zustände von einem Versuchsleiter erfragen, Antwort entspricht qual. Niveau der Frage; exp. Bed.: mit/ohne Reflexion des Problemlösens in einem anderen Managementspiel
Vent (Denkstrategie 1985), S. 90f., 106, 115, 126, 137f.	volkswirtschaftliches Planspiel: Optimierung der Energieversorgung privater Haushalte der BRD bis 2023, hochkomplexes auf empirischen Daten beruhendes System, Effizienzmaß stützt sich auf 17 Variablen, dabei gehen Kosten mit 50 v. H. des Gesamtgewichtes ein	Rückmeldung über 17 zentrale Systemvariablen und nicht realisierbare Maßnahmen, Aufschlüsselung der Kosten und weitere Informationen können vom Versuchsleiter erfragt werden; exp. Bed.: 1. tabell. vs. graphische Systemdarstellung, 2. tabell. vs. graphisches Feedback

Abbildung 3.6.: Maßgrößen der Prognosegüte in (experimentellen) empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Pankoff/Virgil (Usefulness 1970), S. 9, 13	Beurteilung, wie sich der Aktienkurs am Ende des Jahres entwickeln wird, subjektive Wahrscheinlichkeit der realisierten Antwortmöglichkeit wird als Effizienzmaß verwendet	35 Informationen über gesamtwirtschaftliche Lage, Branche und Unternehmen für je drei Jahre und drei Firmen, kostenpflichtige Beschaffung, historische Daten
Stael von Holstein (Probabilistic Forecasting 1972), S. 141, 143, 153	kurzfristige Aktienkursprognose, Übereinstimmung der subjektiven Wahrscheinlichkeitsverteilung mit tatsächlich realisiertem Ereignisintervall, verschiedene Scoring-Regeln	Feedback über die eigenen und fremden Wahrscheinlichkeitsverteilungen, die zugeordneten Scores und die Rangplätze bei den verschiedenen Sitzungen
Smith (Inquiry Techniques 1975), S. 33	Unternehmensspiel, Prognose von Gesamtkosten und Stückkosten, relative absolute Abweichung des prognostizierten vom tatsächlichen Wert	exp. Bed.: Abfrage von bis zu 17 Berichten vs. zusätzliche individ. aufgebaute Berichte a) tabellarisch, b) in (grober) graphischer Form
Brandon/Jarett (Forecast 1977), S. 699 ff.	Prognose von Kurs-Gewinn-Verhältnissen einer simulierten Unternehmung, absolute Abweichungen der prognostizierten von den simulierten Werten, deren Steigerungsraten und eines Konfidenzintervalles um die Werte	5 experimentelle Angebotsbedingungen mit unterschiedlichem Umfang und Inhalt, außerdem werden die Qualität der angebotenen Prognosen und die Variabilität der zu prognostizierenden Daten variiert
Murphy/Winkler (Forecasts 1977), S. 49 ff.	kurzfristige Temperaturprognose, Vergleich von Median- und Intervallschätzungen mit tatsächlicher Temperatur und klimatologischem Durchschnitt, absolute Abweichung vom Median, Breite des 50 und 75-%-Intervalls	Feldstudie mit Meteorologen, die auf die üblichen Unterlagen und Hilfsmittel ihrer Institutionen zurückgreifen können, Prognosen werden als außerordentlich genau angesehen
Crichfield/Dyckman/Lakanishok (Forecasts 1978), S. 653 ff.	Prognose des Kurs-Gewinn-Verhältnisses ("earnings per share"), die Qualität der Prognosen wird nach Theils U^2 - Statistik bestimmt	alle Informationsquellen, die amerikanischen Finanzanalytikern bei ihren Prognosen zur Verfügung stehen
Ebert/Kruse (Bootstrapping 1978), S. 113	Prognose der erwarteten jährlichen Rendite von Wertpapieren, Korrelation zwischen prognostizierter und tatsächlicher Rendite, Rangkorr. zwischen prognostizierter und tatsächlicher Rangfolge der Wertpapiere	22 Informationen über gesamtwirtschaftliche Lage, Branche und Unternehmen, 35 zu beurteilende Unternehmen
Brockhoff (Delphi-Prognosen 1979), S. 24, 111 ff. Vgl. auch Brockhoff (Forecasting Quality 1984), S. 419 ff.	Prognose von Zinssätzen auf kurze Sicht (drei bzw. sechs Monate), relative absolute Abweichung des prognostizierten vom tatsächlich realisierten Wert	87 verschiedene Zeitreihen, die in unterschiedlicher Präsentationsform aus einer Datenbank abgerufen werden können, außerdem werden die Schätzungen anderer Experten im Rahmen einer via Computerkonferenz realisierten Delphi-Gruppe mitgeteilt

Abbildung 3.6.: Maßgrößen der Prognosegüte in (experimentellen) empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Smith (Segment Reporting 1979), S. 138 f.	Prognose des Betriebserfolges ("operating earnings"), Absolutbetrag des prozentualen Schätzfehlers	vergangenheits- und zukunftsbezogene finanzielle Daten; exp. Bed.: Segmentberichterstattung, ohne, grob, detailliert
Schütt (Wahrscheinlichkeitsschätzungen 1979), S. 45 ff., 246 ff.	kurzfristige Dollarkursprognose, Übereinstimmung der subjektiven Wahrscheinlichkeitsverteilung mit tatsächlich realisiertem Ereignisintervall, verschiedene Scoring-Regeln	Experimental- und Kontrollgruppe, Experimentalgruppe wird durch ein interaktives Entscheidungsunterstützungsverfahren im Computer-Dialog unterstützt
Wright (Models 1979), S. 75 f., Wright (Models 1977), S. 680	Prognose der erwarteten Kursveränderung einer Aktie von einem Jahr zum nächsten, erfolgsabhängiges Honorar der Versuchspersonen	5 ausgewählte Bilanzkennzahlen und Börseninformationen
Ashton (Predictions 1982), S. 443 ff, S. 447	Prognose der für das ganze Jahr bestellten Anzeigenseiten des Time Magazins, absolute Abweichungen und Korrelationen von prognostizierten und tatsächlichen Seitenzahlen	5 Informationen: Quartal (1,2,3), Anzeigenseiten pro Quartal, Auto- und Spirituosenanzeigenseiten pro Quartal, bis Jahresende gebuchte Anzeigenseiten ("Linsen"-Studie)
McIntyre (Marketing Models 1982), S. 24	simulierte Verkaufseinsatzplanung auf Basis des CALLPLAN-Models von Lodish, Umsatzprognose, absolute Abweichung des geschätzten vom tatsächlichen Umsatz, dividiert durch den absoluten Umsatz, exp. konstant gehaltene Umweltbedingungen mit stochast. Fehlerterm	Vpn müssen subj. Reaktionsfunktionen schätzen und Budgetaufteilung auf versch. Verkaufsgebiete vornehmen, (vorher gleich) Feedback über erzielte Gewinne und Umsätze der drei bzw. sechs Verkaufsgebiete
Otley/Dias (Aggregation 1982), S. 174, 176 f., 178 f.	simulierte Produktprogrammplanung, Schätzen von Produktdeckungsbeiträgen und Gewinn, prozentualer Schätzfehler, bei exp. konstant gehaltenen Umweltbedingungen	unterschiedlich ausführliche Berichte über Produktion, Absatz und Erfolg, vier exp. Aggregationsbedingungen
Pratt (Post-Cognitive Structure 1982), S. 198	Prognose des Netto-Einkommens einer Unternehmung, quadratische Abweichung von prognostiziertem und tatsächlichem Wert	Jahresabschlüsse großer amerikanischer Unternehmen; exp. Bed.: Komplexität der Berichte, 3 Klassen lt. Expertenrating
Ashton (Field Test 1984), S. 365, vgl. auch Ashton/Ashton (Aggregating Subjective Forecasts 1985), S. 1501 ff.	Prognose der vierteljährlich bestellten Anzeigenseiten des Time Magazins, absolute und prozentuale Abweichungen der prognostizierten von den tatsächlichen Seitenzahlen	Feldstudie von Top-Managern, alle Informationsquellen der realen Entscheidung

Abbildung 3.7.: Maßgrößen der Angemessenheit von Gesamturteilen in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Barefield (Aggregation 1972), S. 237	Analyse von Kostenabweichungen, Beurteilung, ob Arbeitskräfte wirtschaftlich eingesetzt werden oder nicht, Differenz von tatsächlich korrekt klassifizierten Umweltzuständen und Anzahl korrekt klassifizierbarer Umweltzustände auf Basis eines statistischen Modells	verschiedene Kostenabweichungen, Feedback während der Trainingsphase, ob richtig klassifiziert wurde (40 Fälle), weitere 40 Testfälle zur Bestimmung der Effizienz, exp. Bed.: Aggregationsgrad der berichteten Abweichungen und Korrelation der Ursachen
Abdel-khalik (Aggregating Reports 1973), S. 131	Beurteilung, ob ein Firmenkunde Tilgungen und Zinszahlungen einstellt, Verhältnis der geschätzten Kreditausfallwahrscheinlichkeiten von solventen und insolventen Firmenkunden	drei exp. Bed.: bilanzielle Urdaten, teilw. verdichtete Daten, auf 12 Summenpos. verdichtete Daten, je drei Jahre, vier zu beurteilende Unternehmen
McIntyre (Financial Statements 1973), S. 578, 580 f.	Aktienkauf, Beurteilung, welche von zwei angebotenen Firmen im Verlauf des nächsten bzw. der nächsten zwei, drei und mehr Jahre die höchste Rendite erbringen wird, Vergleich mit tatsächlicher historischer Entwicklung	exp. Bed.: Jahresabschlüsse nur auf Basis von historischen Kosten vs. nur auf Basis von Wiederbeschaffungskosten, beide Informationen, drei zurückliegende Jahre, je zwei zu beurteilende Unternehmen
Kennedy (Usefulness 1975), S. 100, 103, 107	Schätzen der Wahrscheinlichkeit, daß ein Unternehmen im nächsten Jahr insolvent wird, Revision der subjektiven Wahrscheinlichkeit nach jeder erhaltenen Kennzahl, Prozentsatz der Fälle, in denen die Revision in die richtige Richtung erfolgt	Branchenzugehörigkeit, Bilanzsumme, vier Bilanzkennzahlen, jeweils für das Jahr vor Insolvenz, 12 zu beurteilende Unternehmen
Libby (Decision Makers 1975), S. 481	Beurteilung, ob ein Unternehmen innerhalb der nächsten drei Jahre insolvent wird oder nicht, Anzahl der korrekt klassifizierten Fälle	fünf Bilanzkennzahlen für eines der drei Jahre vor Insolvenz, 60 zu beurteilende Unternehmen
Vasarhelyi (Planning Systems 1977), S. 145	simulierte Planungsentscheidung, Qualität der Entscheidung nach Expertenurteil bewertet	Daten- und Programmbank mit zahlreichen Informationen, ca. 80 APL-Programme
Baxendale (Decision Accuracy 1978), S. 80	Analyse von Kostenabweichungen, Beurteilung, ob Arbeitskräfte wirtschaftlich eingesetzt werden oder nicht, Anzahl korrekt klassifizierter Umweltzustände	verschiedene Kostenabweichungen, Feedback während der Trainingsphase, ob richtig klassifiziert wurde (40 Fälle), weitere 40 Testfälle zur Bestimmung der Effizienz
Brown (Information Processing 1978), S. 43 f.	Analyse von Kostenabweichungen, Beurteilung, ob ein Produktionsprozeß "normal" verläuft oder nicht, Gesamtkosten der Fehlklassifikationen für Fehler erster und zweiter Art, im Vergleich zu einem stat. optimalen Modell	exp. Bed.: ohne/mit statistischen Zusatzinformationen; Kostenverhältnisse für Fehler erster und zweiter Art; Überlappung der Verteilung der Regelgröße bei beiden Zuständen des Prozesses: klein/groß
Elstein/Shulman/Sprafka (Problem Solving 1978), S. 60	Diagnose von Krankheiten, Beurteilung, ob die nach Expertenmeinung richtige(n) Diagnose(n) gestellt wird (werden)	Gespräch mit dem Patienten (von Schauspielern simuliert), Erfragen von Laborbefunden von einer Datenbank
Moriarty (Graphics 1979), S. 212, 214	Beurteilung, ob ein Unternehmen insolvent wird oder nicht, Anzahl der falsch klassifizierten Fälle, Honorar der Versuchspersonen hängt von der Trefferzahl und dem Zeitaufwand ab, außerdem wird der Quotient aus Trefferzahl und Zeitaufwand als Effizienzmaß verwendet	vier exp. Bed.: schematische Gesichter a) ohne, b) mit Erklärung, c) 10 ausgewählte Daten aus Bilanz und Erfolgsrechnung, d) 13 ausgewählte Bilanzkennzahlen; jeweils für sechs Jahre vor dem Insolvenzjahr, 22 Unternehmen

Abbildung 3.7.: Maßgrößen der Angemessenheit von Gesamturteilen in experimentellen empirischen Studien

Quelle	Effizienzmaß	Informationsbasis
Abdel-khalik/EI Sheshai (Information Choice 1980), S. 329	Beurteilung, ob ein Firmenkunde Tilgungen und Zinszahlungen einstellt ("default on loans"), Anzahl der korrekten Klassifikationen	ausgewählte Kennzahlen, gebührenpflichtige Nachfrage (max. 8 Kennzahlen), 32 zu beurteilende Unternehmen
Casey (Information Load 1980), S. 40, 44 vgl. auch Casey (Usefulness 1980), S. 609	Beurteilung, ob eine Unternehmung innerhalb der nächsten drei Jahre insolvent wird, Anzahl der korrekten Klassifikationen (von 9 bzw. 10 ausgewählten Fällen)	3 exp. Bed.: a) sechs ausgewählte Kennzahlen, b) zusätzlich Bilanz und Erfolgsrechnung, c) zusätzlich Erläuterungen; jeweils fünftes bis drittes Jahr vor der Insolvenz, 30 zu beurteilende Unternehmen
Friedman/Neumann (Opportunity Costs 1980), S. 410, 413 f.	Beurteilung, welches von zwei vorgegebenen Investitionsprojekten unter Berücksichtigung von Opportunitätskosten günstiger ist	vier Kosteninformationen, aus fünf weiteren können max. 2 zusätzlich beschafft werden, 12 zu beurteilende Projektpaare
Kida (Judgments 1980), S. 511	Beurteilung von Wirtschaftsprüfern, ob mit "going-concern"-Problemen zu rechnen ist, Anzahl der korrekt klassifizierten Fälle	fünf Kennzahlen für das Jahr, vor dem die Unternehmung die Definition von "going-concern"-Problemen erfüllt, 40 Unternehmen
Zimmer (Lens Study 1980), S. 631	Beurteilung, ob ein Unternehmen innerhalb des nächsten Jahres insolvent wird oder nicht, Anzahl der korrekt klassifizierten Fälle	fünf Bilanzkennzahlen für drei Jahre vor der Insolvenz, 42 zu beurteilende Unternehmen
Kessler/Ashton (Feedback 1981), S. 152 f.	Bond-Rating, Übereinstimmung des abgegebenen Ratings mit dem von Moody's (Korrelationskoeffizient und absolute Abweichung)	drei Kennzahlen für ein Jahr; exp. Bed.: Feedback über Trefferquote vs. Feedback über Zusammenhang von Kennzahlen und Ratings (versch. Varianten), 34 Bonds
Casey (Probability Disclosure 1983) S. 304	Beurteilung, ob eine Unternehmung innerhalb der nächsten beiden Jahre insolvent wird, Anzahl der korrekten Klassifikationen (von 49 Fällen, davon 15 insolvente, 34 nicht insolvente)	Branche, sechs ausgewählte Kennzahlen mit Branchendurchschnitten für die letzten beiden Jahre, Gesamtvermögen, Trends für für drei Kennzahlen für die letzten drei Jahre, 49 zu beurteilende Unternehmen, exp. Bed.: Anteil der Insolvenzen in der Grundgesamtheit bekanntgegeben ja/nein
Shields (Supply and Demand 1983), S. 291 f.	Analyse von Kosten- und Leistungsberichten betrieblicher Teileinheiten "performance reports", Angemessenheit des Urteils: ob die Analytiker die nach Expertenmeinung "kritische Ursache" herausgefunden haben	exp. Bed.: 3 vs. 9 zu beurteilende Teileinheiten bzw. Berichte mit 6 vs. 13 Informationseinheiten, die Daten wurden von information-display-boards beschafft
Stöck/Watson (Graphics 1984), S. 195 f., 198 f., 203 f.	Bond-Rating, Beurteilung, ob das von Experten abgegebene Rating im nachfolgenden siebten Jahre steigt, fällt oder gleich bleibt, Anteil korr. Klassifikationen, erfolgsabh. Honorar	sechs Kennzahlen für sechs Jahre; exp. Bed.: a) Kennzahlen, b) schematische Gesichter, c) Gesichter und Wahrscheinlichkeiten der Diskriminanzanalyse
Chalos (Financial Distress 1985), S. 535, 541	Beurteilung, ob ein Unternehmen innerhalb des nächsten Jahres insolvent wird oder nicht, Anzahl der korrekt klassifizierten Fälle	Bilanzdaten, 8 Kennzahlen und Branchendurchschnitte für drei Jahre vor der Insolvenz, 8 zu beurteilende Unternehmen

3.2. OPERATIONALISIERUNG DER TREFFSICHERHEIT DER LAGEBEURTEILUNG

3.2.1. Das Meßziel

Die Effizienzdimension "Treffsicherheit der Lagebeurteilung" stützt sich auf die von den Analytikern gebildeten Teilurteile. Zur Bestimmung des Analysenutzens dieser Teilurteile für die Vorbereitung des Entscheidungsprozesses werden drei Kriterien herangezogen:

- a) die Relevanz der beurteilten Tatbestände ("was wird beurteilt"),
- b) die Qualität der Ableitung der Urteile ("wie wird beurteilt"),
- c) die Menge der beurteilten Tatbestände ("wieviel wird beurteilt").

Zu a) Relevanz der beurteilten Tatbestände

Eine valide Effizienzmessung verlangt, daß man sich auf anerkannte Bewertungsmaßstäbe stützt. Da es wohl kaum "den" Beurteilungsmaßstab für "die" Bilanzanalyse gibt, legen wir unserer Effizienzmessung **mehrere** Bewertungsmaßstäbe zugrunde, die man als angemessen ansehen kann. Diese Maßstäbe werden objektiviert und einheitlich angewandt. Außerdem wird geprüft, wie stark die Art des gewählten Maßstabes die Leistungsbeurteilung beeinflußt.

Unsere Relevanzbestimmung stützt sich auf die folgenden drei Relevanzmaßstäbe:

1. Das **theoretische Wissen bilanzanalytischer Lehrbücher**: Hierzu wird eine systematische Inhaltsanalyse der bilanzanalytischen Monographien durchgeführt, um zu bestimmen, welche Tatbestände nach Meinung ihrer Autoren, die man als Experten ansehen kann, zu beurteilen sind. Als Relevanzmaßstab wird die Anzahl der **Kennzahlenvorschläge**, die zu einem Urteilsbereich unterbreitet werden, verwendet.

2. Das **empirische Wissen statistischer Untersuchungen**: Hierzu wird eine eigene statistische Auswertung von krisenfreien und krisenbehafteten Unternehmen durchgeführt. Als Relevanzmaß wird der **Anteil richtig klassifizierter Unternehmen** bezüglich einer repräsentativen Kennzahl, die einem Urteilsbereich zugeordnet ist, verwendet. Um das Gebot der Angemessenheit der Effizienzmessung zu wahren, bezieht sich diese Auswertung auf die gleiche Basis wie unsere Experimente. Um die Validität des Relevanzmaßstabes zu erhöhen, werden die Klassifikationsgüten getrennt für Norm-, Zeit- und Betriebsvergleiche bestimmt. Diese Vergleichsarten diskriminieren nämlich erfolglose und erfolgreiche Unternehmen unterschiedlich gut, und sie werden von den Analytikern in verschiedenartiger Weise eingesetzt.

3. Die **Bedeutsamkeit des Urteilsbereiches** für den konkreten Fall: Zur Bestimmung der fallspezifischen Relevanz eines Urteilsbereiches wird geprüft, wie stark seine fallspezifische Ausprägung von der mittleren Ausprägung des Samples abweicht. Als Relevanzmaßstab wird der sogenannte "z-Wert" der repräsentativen Kennzahl, die einem Urteilsbereich zugeordnet ist, verwendet. Man erhält diesen **z-Wert**, indem man von der Kennzahl des analysierten Falles die mittlere Kennzahl ausprägung des Firmensamples abzieht und diese Differenz durch die Standardabweichung der Kennzahl im untersuchten Firmensample dividiert.¹⁾ Auch diese Relevanzmaße werden nach den Vergleichsarten differenziert: Eine Firma kann im Betriebsvergleich noch relativ gut aussehen, weil sie und ihr Vergleichsfall einer Branche angehören, die in einer Krise steckt. Im Zeitvergleich ergibt sich jedoch für beide Unternehmen eine ungünstige Entwicklung.

Zu b) Qualität der Ableitung der Teilurteile

Für die Effizienzdimension Treffsicherheit der Lagebeurteilung sollen nur solche Urteile verwendet werden, die als "sachgerecht" anzusehen sind. Zur Objektivierung dieses Qualitätsmerkmals werden folgende Anforderungen formuliert:

1. Das Urteil muß auf einer **faktisch wahren** Aussage beruhen. Wenn beispielsweise behauptet wird, daß der Umsatzanstieg erfreulich sei, de facto aber ein Umsatzrückgang vorliegt, dann handelt es sich um ein "nicht sachgerechtes" Urteil zur Umsatzentwicklung.
2. Das Urteil muß auf **logisch wahren** Schlüssen beruhen. Wenn behauptet wird, daß der Rückgang des Umlaufvermögens darauf beruhe, daß die Beteiligungen gesunken sind, dann wird diese Aussage als "nicht sachgerecht" beurteilt, weil die Beteiligungen keinen Bestandteil des Umlaufvermögens darstellen. Die Aussage ist schon aus "definitions"-logischen Gründen als falsch anzusehen.
3. Das Urteil muß sich auf eine **angemessene Bewertungsrichtung** stützen. Wenn ein starker Rückgang des Betriebserfolges als "positiv" angesehen wird, ohne daß besondere Gründe des Einzelfalls angegeben werden, die diese Einschätzung rechtfertigen, dann wird dies als ein "nicht sachgerechtes" Urteil gewertet. Diese Anforderung wird allerdings nur bei jenen Positionen gestellt, bei denen man eine generell angemessene Bewertungsrichtung feststellen kann. Dies gilt im Prinzip nur bei Aussagen zur Erfolgslage. Ausprägungen und Veränderungen von Vermögens- und Kapitalpositionen lassen sich im allgemeinen nicht generell als positiv oder negativ einschätzen. So kann z. B. ein hoher Anteil des Anlagevermögens an der Bilanzsumme sowohl positive als auch negative Tatbestände anzeigen.
4. Das Urteil muß auf einer **sachgerechten Relativierung** beruhen. So wird bei Betriebsvergleichen unterschiedlich großer Unternehmen stets verlangt, daß die verglichenen Daten an einem geeigneten Größenmaßstab relativiert werden.
5. Das Urteil muß sich auf **angemessen gebildete Kennzahlen** stützen. Was "angemessen" ist, hängt vom **Meßziel** des jeweiligen Urteilsbereiches ab, dem eine Aussage zugeordnet wird. Zur Objektivierung dieser Einschätzung wurden in mehrwöchigen Gruppendiskussionen **Bewertungstabellen** erarbeitet, in denen die für einen Urteilsbereich in Frage kommenden Kennzahlen als "nicht", "teilweise" oder "völlig" angemessen eingestuft werden. So

wird beispielsweise zum Bereich "120 Qualifizierende Aussagen zum Gesamterfolg mit Hilfe von Renditemaßen" eine Tabelle mit 84 Elementen vorgegeben (14 Zähler- und 6 Nennergrößen)²⁾. Danach sollte im Zähler nicht der Bilanzgewinn, sondern wenigstens der Jahresüberschuß stehen. Im Nenner sollten das Gesamtkapital, die Gesamtleistung oder der Umsatz herangezogen werden.

Zu c) Menge der beurteilten Tatbestände

Ein Analytiker, der mehr Tatbestände beurteilt als ein anderer ist als effizienter einzustufen, weil er umsichtiger vorgeht und auf mehr Stärken und Schwächen eines Falls hinweist und damit die Lage eines Unternehmens umfassender würdigt. Wir berücksichtigen diese Überlegung, indem wir die Treffsicherheit der Lagebeurteilung als eine gewichtete Summe der Meßwerte der Teilurteile bilden:

$$\text{Treffsicherheit der Lagebeurteilung} = \sum_i U_i * W_i$$

Hierbei bedeuten:

i : Laufindex für die Urteilsbereiche, $i = 1, \dots, I$

U_i : Bool'sche Variable, die den Wert Eins hat, falls zum Urteilsbereich i ein qualifiziertes Urteil abgegeben wird. Andernfalls hat die Variable den Wert Null.

W_i : Relevanzgewicht des Urteilsbereiches i

Diese Formel wird im Verlauf der Arbeit noch differenziert.

Gegen diese Operationalisierung läßt sich einwenden, daß sich bei sehr vielen abgegebenen Urteilen beim Adressaten wegen dessen begrenzter Aufnahme- und Verarbeitungskapazität dysfunktionale Effekte ergeben können. Außerdem wird nicht berücksichtigt, daß zwischen den Aussagen zu den ein-

zelnen Urteilsbereichen Redundanzen bestehen können.

Diese Einwände lassen sich entkräften, indem man statt einer linearen Aufsummation der Teilnutzenwerte der einzelnen Urteilsbereiche komplexere Formeln entwickelt. Das Problem besteht jedoch darin, daß man zur **Anwendung** dieser Formeln schwer zu beschaffendes empirisches Wissen benötigt:

- * Man muß genau wissen, wie die Entscheidungseffizienz eines repräsentativen Adressaten in Abhängigkeit von unserem Meßwert verläuft, um eine genau kalibrierte Nutzenfunktion zu konstruieren, die diesen Einfluß berücksichtigt. Hierzu müßte man einem repräsentativen Sample von Analytikern zu einem ebenfalls repräsentativen Sample von Kreditfällen je einen Satz von Bilanzanalysen vorlegen, die sich über einen hinreichenden Wertebereich bezüglich der von uns skalierten Treffsicherheit der Lagebeurteilung unterscheiden. Dann müßte geprüft werden, welche Beziehung zwischen der Qualität der verwendeten Bilanzanalyse und der Güte der Entscheidung besteht.

- * Ähnlich komplex ist die Bestimmung der Redundanzen. Im Prinzip müßte für jedes Tupel von bereits sachgerecht beurteilten Bereichen geprüft werden, welche Analysenutzen ein bestimmter, zusätzlich beurteilter Bereich stiftet. Selbst wenn der Analysezweck nur darin bestünde, einen Fall als "mutmaßlich solvent" oder als "mutmaßlich insolvent" zu klassifizieren, würde der Rechenaufwand zur Bestimmung der Redundanzfunktionen rasch astronomische Größenordnungen erreichen.

Man muß daher kritisch fragen, ob sich der Aufwand für eine derartige Verfeinerung lohnt. Bei unserem Meßkonzept haben naive "Vielschreiber" nur eine geringe Chance hohe Treffsicherheitswerte zu erreichen, denn:

- * nur sachgerecht abgeleitete Urteile werden berücksichtigt,
- * wichtige Urteilsbereiche gehen mit einem höheren Gewicht in die Formel ein als unwichtige und

* für die Analysen steht nur ein begrenzter Zeitraum von 90 Minuten zur Verfügung.

Die weiter unten dargestellten empirischen Befunde bestätigen diese Überlegungen: Auch gut bewertete Analytiker beurteilen nur einen Teil der möglichen Bereiche. Dies liegt daran, daß nicht nur die Adressaten der Bilanzanalyse, sondern auch die Analytiker eine begrenzte Kapazität haben, um Tatbestände zu prüfen und sachgerecht zu beurteilen. Auch wenn zwischen unserem Summenmaß und dem subjektiven Entscheidungsnutzen für einen typischen Adressaten ein umgekehrt u-förmiger Zusammenhang besteht, können wir doch mit großer Sicherheit davon ausgehen, daß die hier gemessenen Summenwerte auf dem aufsteigenden Ast der Kurve liegen und zwar so, daß man sie gut durch eine Gerade approximieren kann. Wir unterstellen deshalb eine Proportionalität zwischen dem Effizienzmaß "Treffsicherheit der Lagebeurteilung" und dem Entscheidungsnutzen eines repräsentativen Adressaten der zu bewertenden Bilanzanalyse.

Man sollte die Frage der Zusammenfassung der Teilurteile zu einem Gesamturteil nicht unangemessen stark gewichten: Ein auch bei betriebswirtschaftlichen Anwendungen häufig bestätigtes Ergebnis der bereits kritisch gewürdigten Linsen-Modell-Studien, in deren Mittelpunkt die Aggregation von Einzelurteilen zu einem Gesamturteil steht, lautet, daß man das Gesamturteil mit einer außerordentlich hohen Genauigkeit aus einer gewichteten Summe der Teilurteile prognostizieren kann.³⁾ Deshalb ist es unseres Erachtens angemessen, wenn wir der gängigen Praxis von Scoring-Modellen folgen und auch bei unserem Effizienzmaß eine lineare Gewichtung der Teilurteile vornehmen.⁴⁾

Substantiell beachtlicher als die intensiv erforschte Frage der Zusammenfassung der Teilurteile zu einem Gesamturteil ist unseres Erachtens die Bestimmung der Relevanzmaße für die Teilurteile und die Operationalisierung der Anforderungen an die Qualität der Urteilsbildung. Abbildung 3.8. faßt die hierzu erarbeiteten Anforderungen zusammen. Mit ihrer Umsetzung befassen sich die nachfolgenden Kapitel. Zuvor ist das System der Urteilsbereiche zu entwickeln, auf das sich unsere Effizienzmessung bezieht.

Abbildung 3.8.: Anforderungen an die Effizienzdimension Treffsicherheit der Lagebeurteilung

Anforderungen an die Ableitung
der gebildeten Urteile

Faktisch wahre Prämissen

Logisch wahre Schlüsse

Pragmatisch angemessene
Bewertungsrichtung

Angemessene Relativierung
von Größeneinflüssen

Sachlich angemessene
Kennzahlenbildung

Treffsicherheit
der
Lagebeurteilung

Anforderungen an die Relevanz der
beurteilten Tatbestände

"Belesenheit" – gemessen am theoretischen
Wissen bilanzanalytischer Lehrbücher

"Erfahrenheit" – gemessen am empirischen
Wissen um statistisch ermittelte Unter-
schiede krisenfreier und krisenbehafteter Unternehmen

"Aufmerksamkeit" – gemessen an der Fähig-
keit, fallspezifische Auffälligkeiten
aufzudecken

3.2.2. Die Bestimmung der Urteilsbereiche

Nach welchen Kriterien soll man die Urteile gliedern? Wieviele Urteilsbereiche soll man unterscheiden? Auf diese Fragen lassen sich zwar keine eindeutigen Antworten geben, aber es können einige Optimierungsüberlegungen zur Anzahl der Urteilsbereiche angestellt werden.

Je mehr Urteilsbereiche gebildet werden, desto genauer kann man die Aussageziele abgrenzen. Man kann einem Bereich spezifischere Kennzahlen zuordnen und die Relevanz von Urteilen, die zu diesem Bereich gebildet werden, feiner abstufen. Der Soll-Ist-Vergleich von theoretisch vorgesehenen und empirisch beurteilten Tatbeständen wird präziser, weil schärfer geprüft werden kann, ob ein getroffenes Ist-Urteil den besonderen Anforderungen eines Urteilsbereiches entspricht.

Auf der anderen Seite wird es mit zunehmender Kategorienzahl für einen Erheber schwieriger, ein Ist-Urteil einem Bereich zuzuordnen, weil mehr und sich ähnlicher werdende Urteilsbereiche für eine Codierung in Frage kommen. Den höheren Codierungszeiten stehen sinkende Grenzerträge der gebildeten Skalen gegenüber: Je mehr Urteilsbereiche vorgesehen werden, desto häufiger wird es vorkommen, daß ein Analytiker zu einem Bereich keine Aussage trifft. Die Bereiche tragen daher immer weniger zur interpersonellen Differenzierung der Analysequalität bei.

Wir stehen somit vor einem vielschichtigen **Adäquanzproblem**.⁵⁾ Unser Lösungsvorschlag sieht eine **hierarchische Systematisierung der Urteilsbereiche** vor. Sie hat den Vorteil, daß man bei der **Erhebung** der Ist-Aussagen den Ermessensspielraum schrittweise verringern kann, indem man immer feiner bestimmten Ästen des Klassifikationsbaumes folgt. Bei der **Auswertung** besteht die Möglichkeit, zunächst einmal die zuverlässiger gemessenen Urteile auf der höchsten Aggregationsstufe zu betrachten und dann eine fortschreitend feinere Zerlegung vorzunehmen.

Es stellt sich nun die Frage, nach welchen Kriterien man die Urteilsbereiche abgrenzen soll. Grundlegend erscheint uns die Trennung des Jahresabschlusses in Bilanz und Gewinn- und Verlustrechnung. Dieser Einteilung

Es stellt sich nun die Frage, nach welchen Kriterien man die Urteilsbereiche abgrenzen soll. Grundlegend erscheint uns die Trennung des Jahresabschlusses in Bilanz und Gewinn- und Verlustrechnung. Dieser Einteilung folgen die in der Wirtschaftspraxis vorgesehenen Aufbereitungsformulare,⁶⁾ zahlreiche Lehrbücher zur Bilanzanalyse⁷⁾ und das in unseren Experimenten verwendete Angebot an aktienrechtlichen Primärinformationen und aufbereiteten Daten. Schließlich zeigen auch die Befunde von Weigel⁸⁾ und Knorr,⁹⁾ daß sowohl Kreditmanager als auch studentische Analytiker den Verlauf ihres Analyseprozesses an dieser Grundgliederung ausrichten. Es erscheint daher sinnvoll, zwischen Urteilen zur "Erfolgslage", und Urteilen zur "Vermögens- und Finanzlage" zu unterscheiden.

Fraglich ist, ob wir überdies zwischen Urteilen zur "Vermögenslage" und Urteilen zur "Finanzlage" differenzieren sollten. Zur Abgabe von Urteilen über die "Finanzlage" im Sinne der angestrebten neuen Rechnungslegung würde nämlich ein eigenständiges Rechenwerk benötigt.¹⁰⁾ Unseren Versuchspersonen wurde aber keine "Finanzierungsbeitragsrechnung" oder eine andere Variante einer "Kapitalflußrechnung" angeboten.¹¹⁾ Für die eigene Erstellung einer solchen Rechnung fehlte ihnen offenkundig die Zeit, möglicherweise auch das erforderliche Wissen, denn solche Rechnungen wurden nur sehr selten und auch dann nur unvollständig angefertigt.

Wenn wir von Urteilen über die "Finanzlage" sprechen, dann sind damit vor allem **bilanzbezogene Urteile**¹²⁾ über Ausmaß, Struktur und Entwicklung von Fremd- und Eigenkapital sowie über horizontale Deckungsrelationen gemeint. Lediglich Urteile über finanzwirtschaftlich konzipierte "Cash-Flow"-Kennzahlen¹³⁾ gehen in die Richtung einer zahlungsstromorientierten Betrachtung. Andererseits besitzen die sog. "**Finanzierungsregeln**"¹⁴⁾ in Kreditvergabeentscheidungen noch immer einen so hohen Stellenwert, daß es gerechtfertigt erscheint, sie von den weniger bedeutsamen Urteilen über die "Vermögenslage" abzugrenzen. Zur "Vermögenslage"¹⁵⁾ rechnen wir alle Urteile über Ausmaß, Struktur und Veränderungen der Aktiva und über das heterogen operationalisierte "Investitionsverhalten".

Um die Validität unseres Meßkonzeptes zu sichern, sind wir auch auf den weiteren Gliederungsstufen den Vorschlägen der bilanzanalytischen Literatur gefolgt. So können sich die Urteile zur Erfolgslage auf den "Gesamterfolg" oder auf Teile des Gesamterfolgs beziehen. Dabei kann es sich um Komponenten der Erfolgsverwendung oder der Erfolgsherkunft handeln. Bei der Erfolgsherkunft wird danach differenziert, ob die grob gegliederten Komponenten "Ordentlicher Erfolg" oder "Außerordentlicher Erfolg" oder die feineren Erfolgsklassen "Betriebserfolg", "Finanzerfolg und sonstige laufenden Erträge", "Liquidations- und a. o. Erträge" und "Bewertungsertrag" oder Teilmengen dieser vier Erfolgsklassen beurteilt werden.¹⁶⁾ Insgesamt ergeben sich auf diese Weise 67 verschiedene Urteilsbereiche zur Erfolgs-, Vermögens- und Finanzlage. Die nachfolgenden Abbildungen 3.9. bis 3.11. stellen die Systematik unserer Bereichsgliederung dar. (Der Übersichtlichkeit halber werden nur die oberen Gliederungsebenen dargestellt).

3.2.3. Die Bestimmung der Relevanzmaße

3.2.3.1. Literarische Relevanz

Praktische Erfahrung schlägt sich in Lehrbüchern zur Bilanzanalyse nieder, insbesondere in solchen, die "aus der Praxis für die Praxis" geschrieben werden. Auf der anderen Seite werden in den Werken von stärker "theoretisch" ausgerichteten Forschern die Meß- und Aussageprobleme gelegentlich differenzierter vorgetragen. Es lag daher nahe, einen Querschnitt unterschiedlicher Schriften zu verwenden.

Wir haben insgesamt 24 verschiedene bilanzanalytische Monographien von "Praktikern" und "Theoretikern" ausgewertet und ausgezählt, wieviele **Kennzahlen** sie zu jedem unserer Urteilsbereiche vorsehen.¹⁷⁾ Hinter dieser Operationalisierung steckt die Überlegung, daß ein Urteilsbereich um so wichtiger ist, je mehr Experten eine Kennzahlenbildung zu diesem Bereich vorschlagen und ein Teilurteil empfehlen. Da die Bilanzanalyse eine vergleichsweise gut strukturierte Aufgabe darstellt, rechnen wir mit einem hohen Konsens der Experten,¹⁸⁾ so daß man zugleich Anhaltspunkte für den Informationsbedarf der Bilanzanalyse gewinnt.¹⁹⁾ Daß der solchermaßen be-

Abbildung 3.9.: Gliederung der Urteilsbereiche zur Erfolgslage

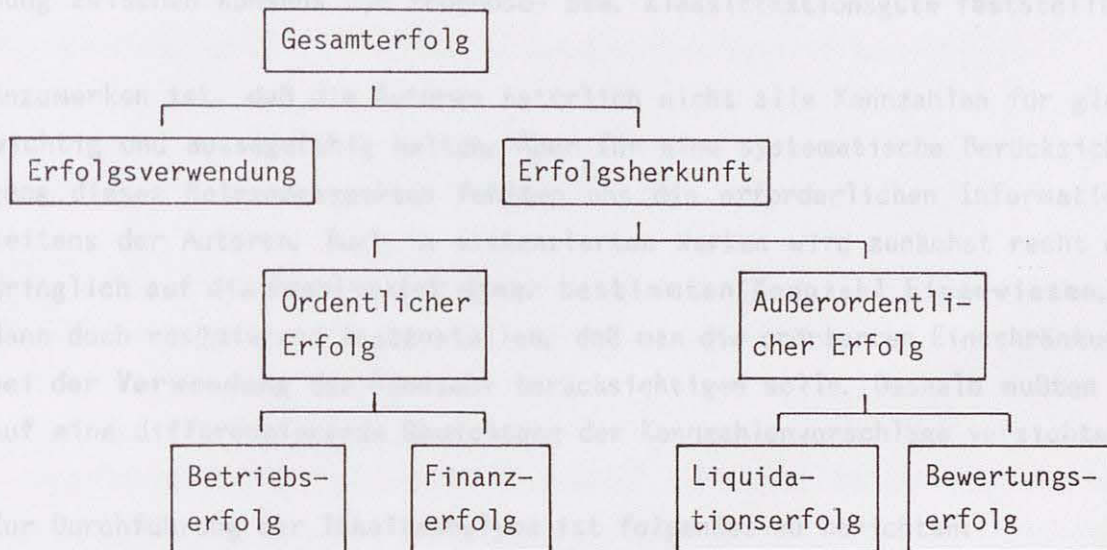


Abbildung 3.10.: Gliederung der Urteilsbereiche zur Vermögenslage

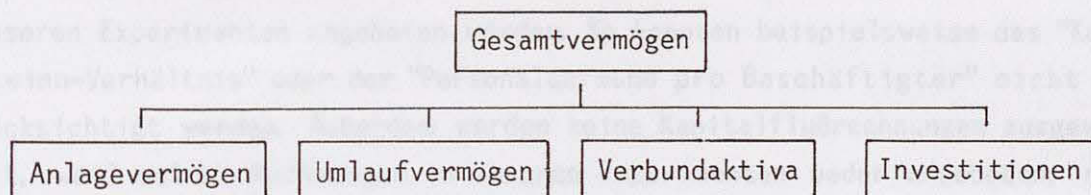
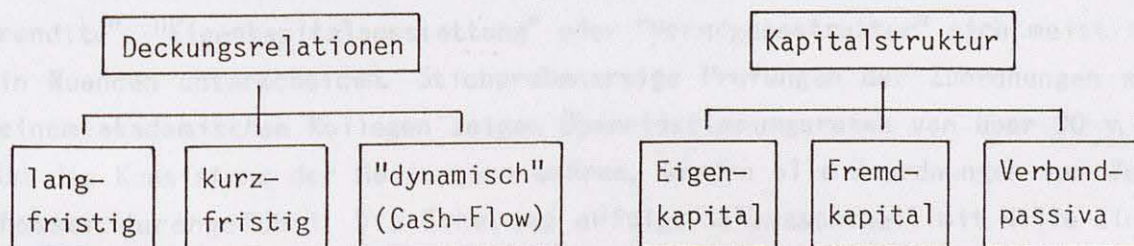


Abbildung 3.11.: Gliederung der Urteilsbereiche zur Finanzlage



stimmte Informationsbedarf für die Problemlösung beachtlich ist, wird vor allem durch die Studie von Ashton belegt, die für zwei vergleichsweise gut strukturierte Beurteilungs- und Prognoseprobleme eine enge positive Beziehung zwischen Konsens und Prognose- bzw. Klassifikationsgüte feststellt.²⁰⁾

Anzumerken ist, daß die Autoren natürlich nicht alle Kennzahlen für gleich wichtig und aussagefähig halten. Aber für eine systematische Berücksichtigung dieses Relevanzaspektes fehlten uns die erforderlichen Informationen seitens der Autoren. Auch in distanzierten Werken wird zunächst recht eindringlich auf die Problematik einer bestimmten Kennzahl hingewiesen, um dann doch resümierend festzustellen, daß man die erörterten Einschränkungen bei der **Verwendung** der Kennzahl berücksichtigen solle. Deshalb mußten wir auf eine differenzierende Gewichtung der Kennzahlenvorschläge verzichten.

Zur Durchführung der Inhaltsanalyse ist folgendes zu berichten:

a) Abgrenzung der ausgewerteten Kennzahlen

Berücksichtigt werden nur Kennzahlen, die sich ausschließlich auf Daten der Rechenwerke Bilanz und Gewinn- und Verlustrechnung des Mindestgliederungsschemas gemäß §§ 151 und 157 Aktiengesetz beziehen, da nur diese Daten in unseren Experimenten angeboten werden. So konnten beispielsweise das "Kurs-Gewinn-Verhältnis" oder der "Personalaufwand pro Beschäftigter" nicht berücksichtigt werden. Außerdem werden keine Kapitalflußrechnungen ausgewertet, weil solche Rechnungen in unseren Experimenten weder angeboten, noch von den Analytikern angefertigt werden.

b) Reliabilität der Messung

Die Zuordnung der Kennzahlen zu den Urteilsbereichen ist relativ unproblematisch, weil die zahlreichen Vorschläge zu "Eigen-" oder "Gesamtkapitalrendite", "Eigenkapitalausstattung" oder "Vermögensstruktur" sich meist nur in Nuancen unterscheiden. Stichprobenartige Prüfungen der Zuordnungen mit einem akademischen Kollegen zeigen Übereinstimmungsraten von über 90 v. H. Um die Konsistenz der Messung zu wahren, werden alle Zuordnungen vom Verfasser durchgeführt. Die Erhebung erfolgt halbmaschinell mit Hilfe eines

Datenbanksystems. Im Ergebnis liegt eine Referenzliste vor, aus der die genaue Zuordnung aller 1239 erhobenen Kennzahlen hervorgeht. Diese kann auf Wunsch gerne eingesehen werden.

Die nachfolgenden Tabellen zeigen das **Ergebnis** unserer Auswertung:

1. Zur **Erfolgs- und Finanzlage** werden wesentlich mehr Kennzahlen vorgeschlagen als zur Vermögenslage (38,3 und 36,6 vs. 24,6 v. H.)
2. In allen drei Kennzahlenblöcken dominiert das Streben nach **verdichtenden Informationen** und **Aussonderung irregulärer Einflüsse**.
3. Bei der **Erfolgslage** kommt dem globalen Erfolgsmaßstab "Gesamterfolg", dem "ordentlichen betrieblichen Erfolg" und dessen Komponenten die höchste Bedeutung zu. Neutrale Erfolge werden gelegentlich in verschiedene Klassen zerlegt, aber in keinem Fall wird eine Kennzahl zur weiteren Zerlegung dieser Erfolgsklassen empfohlen.²¹⁾
4. Die Vorschläge zur Beurteilung der **Vermögenslage** richten sich vor allem auf Struktur und Umschlag von Gesamt-, Anlage- und Umlaufvermögen sowie auf die Investitionstätigkeit. Relativ häufig werden Kennzahlen zur Beurteilung der Entwicklung einzelner kritischer Positionen des Umlaufvermögens vorgeschlagen.
5. Zur Beurteilung der **Finanzlage** werden lang- und kurzfristige Deckungsrelationen, auf dem "Cash-Flow" basierende "dynamische Finanzierungsregeln" und Kennzahlen zur Eigenkapitalausstattung empfohlen. Daneben wird den Rücklagen und den Lieferantenverbindlichkeiten eine gewisse Beachtung geschenkt.

Insgesamt zeigt sich ein recht **hoher Konsens** zwischen den verschiedenen Lehrbüchern, so daß wir uns nicht veranlaßt sehen, weitere Quellen auszuwerten. Wenn man mit allen Vorbehalten eine "Popularitätsskala" der am häufigsten genannten Urteilsbereiche aufstellt, dann zeigt sich das in Tabelle 3.5. dargestellte Ergebnis.

Tabelle 3.1.: Literarische Relevanz der Urteilsbereiche, verdichteter Befund

Urteilsbereiche zur Erfolgslage	Anz.	v.H.
I. Gesamterfolg (1)	126	10,2
II. Erfolgsverwendung (2-5)	35	2,8
III. Erfolgsquellen, ord. vs. a.o. (6-7)	63	5,1
IV. Erfolgsquellen, vier Klassen (8-11)	125	10,1
V. Quellen des Betriebserfolges (12-22)	125	10,1
VI. Quellen der neutralen Erfolge (23-35)	0	0,0
Erfolgslage, alle Bereiche (1-35)	474	38,3
Urteilsbereiche zur Vermögenslage	Anz.	v.H.
I. Gesamtvermögen (36-37)	89	7,2
II. Teile des Gesamtvermögens (38-44)	67	5,4
III. Teile des Anlagevermögens (45-49)	12	1,0
IV. Teile des Umlaufvermögens (50-54)	78	6,3
V. Investitionen (55)	59	4,8
Vermögenslage, alle Bereiche (36-55)	305	24,6
Urteilsbereiche zur Finanzlage	Anz.	v.H.
I. Deckungsrelationen (56-57)	158	12,8
II. Cash-Flow-Kennzahlen (58)	117	9,4
III. Teile des Gesamtkapitals (59-61)	131	10,6
IV. Teile des Eigenkapitals (62-64)	22	1,8
V. Teile des Fremdkapitals (65-69)	26	2,1
Finanzlage, alle Bereiche (56-69)	454	36,6
Deskriptive Kennzahlen zur Unternehmensgröße	6	0,5

Tabelle 3.2.: Literarische Relevanz der Urteilsbereiche zur Erfolgslage

Nr.	Urteilsbereich	Anz. v.H.		Anz. v.H.	
1	Gesamterfolg	126	10,2	126	10,2
2	Steuern v. Einkommen, Ertrag, Vermögen	8	0,6		
3	Ausschüttungen	8	0,6		
4	Rücklagenentwicklung	6	0,4		
5	Zinsaufwendungen	13	1,0	35	2,8
6	Ordentlicher Erfolg	26	2,1		
7	Außerordentlicher Erfolg	37	3,0		
8	Betriebserfolg	93	7,5		
9	Finanzerfolg	29	2,3		
10	Liquidationserfolg	1	0,0		
11	Bewertungserfolg	2	0,1	188	15,2
12	Umsatzentwicklung	20	1,6		
13	Materialeinsatz	20	1,6		
14	Personal: gesamter Aufwand	41	3,3		
15	Personal: Löhne und Gehälter	1	0,0		
16	Personal: Soziale Abgaben	0	0,0		
17	Personal: Altersv. u. Unterstützung	1	0,0		
18	Abschreibg.: gesamter Aufwand	29	2,3		
19	Abschreibg.: Sachanlagen	0	0,0		
20	Abschreibg.: Verluste, Wertmindg.	0	0,0		
21	Abschreibg.: Erträge aus Aufl. SoPo	0	0,0		
22	sonstige Einsatzfaktoren	13	1,0	125	10,1
23	Erträge aus Gewinnbeteiligungen	0	0,0		
24	Erträge aus Beteiligungen	0	0,0		
25	Erträge aus and. Finanzanlagen	0	0,0		
26	sonstige Zinsen u. ä. Erträge	0	0,0		
27	sonstige Erträge ord. Anteil	0	0,0		
28	Abschreibungen auf Finanzanlagen	0	0,0		
29	Aufwendungen aus Verlustübernahme	0	0,0	0	0,0
30	Erträge aus Abgang Anlagevermögen	0	0,0		
31	sonst. Ertr. außerord. Anteil	0	0,0		
32	Einstellg. Sonderposten m. RLA	0	0,0	0	0,0
33	Erträge aus Zuschreibung AV	0	0,0		
34	Herabsetzung Pauschal-WB Forderg.	0	0,0		
35	Erträge aus Auflösung Rückstellg.	0	0,0	0	0,0
	ERFOLGSLAGE, alle Bereiche			474	38,3

Tabelle 3.3.: Literarische Relevanz der Urteilsbereiche zur Vermögenslage

Nr.	Urteilsbereich	Anz. v.H.		Anz. v.H.	
36	Gesamtvermögen, Grobstruktur	67	5,4		
37	Gesamtvermögen, Umschlag	22	1,7	89	7,2
38	Anlagevermögen, Grobstruktur	12	1,0		
39	Anlagevermögen, Umschlag	16	1,3		
41	Umlaufvermögen, Grobstruktur	10	0,8		
42	Umlaufvermögen, Umschlag	28	2,3		
44	Verbundaktiva	1	0,0	67	5,4
45	Grundstücke und Gebäude	2	0,1		
46	Maschinen und maschinelle Anlagen	7	0,5		
47	sonstige Sachanlagen, immaterielles AV	2	0,1		
48	Beteiligungen	1	0,0		
49	sonstige Finanzanlagen	0	0,0	12	1,0
50	Materiallager	10	0,8		
51	Waren- und Erzeugnislager	17	1,4		
52	Debitorenhaltung	29	2,3		
53	monetäres Umlaufvermögen	19	1,5		
54	sonstiges Umlaufvermögen	3	0,2	78	6,3
55	Investitionen	59	4,8	59	4,8
	VERMÖGENSLAGE, alle Bereiche			305	24,6

Tabelle 3.4.: Literarische Relevanz der Urteilsbereiche zur Finanzlage

Nr.	Urteilsbereich	Anz. v.H.		Anz. v.H.	
56	langfristige Deckungsrelationen	57	4,6		
57	kurzfristige Deckungsrelationen	101	8,2	158	12,8
58	Cash-Flow ("dynamische Verschuldung")	117	9,4	117	9,4
59	Eigenkapitalausstattung	92	7,4		
60	Fristigkeit der Finanzierung	39	3,1		
61	Verbundpassiva	0	0,0	131	10,6
62	Grundkapital	0	0,0		
63	Rücklagen ("Börsenkurs")	22	1,8		
64	Sonderposten mit Rücklageanteil	0	0,0	22	1,8
65	Pensionsrückstellungen	0	0,0		
66	andere Rückstellungen	0	0,0		
67	Bankenverbindlichkeiten	5	0,4		
68	Kreditorenhaltung	18	1,5		
69	sonstige kfr. Verbindlichkeiten	3	0,2	26	2,1
	FINANZLAGE, alle Bereiche			454	36,6

Anmerkung: Insgesamt sind 1.239 Kennzahlen erfaßt worden. Davon wurden 6 dem Bereich "70 Deskriptive Kennzahlen zur Unternehmensgröße" zugeordnet.

Tabelle 3.5.: "Popularitätsskala" der häufigsten Urteilsbereiche

Rang	Urteilsbereich	Anzahl	v. H.
1.	Gesamterfolg	126	10,16
2.	Cash-Flow	117	9,44
3.	Kfr. Deckungsrelationen	101	8,15
4.	Betriebserfolg	93	7,50
5.	Eigenkapitalausstattung	92	7,42
6.	Gesamtvermögen (Grobstruktur)	67	5,40
7.	Investitionen	59	4,76
8.	Lgfr. Deckungsrelationen	57	4,60
9.	Personalaufwand	41	3,30
10.	Fristigkeit der Finanzierung	39	3,14
11.	Außerordentlicher Erfolg	37	2,98
	Summe	829	66,85

Man könnte also mit elf Urteilen den Informationsbedarf der Bilanzanalyse zu zwei Dritteln befriedigen, wenn man in Anlehnung an Brockhoff die Vereinigungsmenge der Informationsbedürfnisse ausgewählter bilanzanalytischer Lehrbücher als den "Informationsbedarf" der Bilanzanalyse ansieht.²²⁾

Wenn man jedoch berücksichtigt, daß pro Urteilsbereich häufig mehrere Kennzahlen vorgeschlagen werden, pro Kennzahl mehrere Vergleichsarten herangezogen werden können und daß man bei fallspezifisch wichtigen Kennzahlen auch Urteile über zugrundeliegende Primärinformationen treffen wird, ergibt sich eine Menge von ca. 25 bis 35 zu beurteilenden Tatbeständen. Diese Zahl deckt sich mit dem Befund von Weigel und den weiter unten dargestellten Urteilmengen unserer Analytiker.²³⁾ Mögliche Bedenken, die Treffsicherheit der Lagebeurteilung als eine gewichtete Summe von Teilurteilen zu bestimmen, weil "zu viele" Urteile einen Adressaten überlasten könnten, werden durch den vorliegenden Befund beachtlich relativiert.

3.2.3.2. Statistische Relevanz

Die literarische Relevanzbestimmung entspricht zwar dem Grundsatz der Angemessenheit der Effizienzmessung, weil sie studentischen Analytikern die Möglichkeit gibt, durch Belesenheit eine bessere Analyse anzufertigen, aber sie wirft die Frage auf, ob die in der Literatur vorgeschlagenen Kennzahlen tatsächlich geeignet sind, ein empirisch zutreffendes Gesamturteil abzuleiten. So zeigt sich nämlich in Studien zur Insolvenzprognose auf Basis von numerischen Jahresabschlußinformationen immer wieder, daß nur ein kleiner Teil der aus der Literatur übernommenen Kennzahlen geeignet ist, unterschiedlich definierte "Problem"- und Vergleichsfälle frühzeitig und zuverlässig zu diskriminieren.²⁴⁾ Die Kennzahlenvorschläge der Literatur stützen sich offenkundig nicht immer auf eigene praktische Erfahrungen der Autoren. Es wird vielmehr häufig ungeprüft das übernommen, was andere Autoren für richtig halten.

Dies bedeutet nicht, daß jeder Wissenschaftler nur eigene praktischen Erfahrungen seinen Handlungsempfehlungen zugrundelegen dürfe. Wir fordern vielmehr, daß die Vorschläge auf nachvollziehbaren empirischen Untersuchungen beruhen, die gängige methodische Standards erfüllen. Es bietet sich daher an, unser statistisches Relevanzmaß auf den Ergebnissen der bereits vorliegenden empirischen Untersuchungen zur Prognoseeignung von Bilanzkennzahlen aufzubauen.²⁵⁾

Die Befunde dieser Studien sind für unsere Fragestellung jedoch nur bedingt brauchbar,²⁶⁾ da

- * andere Unternehmensentwicklungen prognostiziert werden,
- * Stichproben mit anderen Rechnungslegungsvorschriften ausgewertet werden,
- * die getesteten Kennzahlen nur einen Teil unserer Urteilsbereiche abdecken und nur für einen kleinen Teil der jeweils getesteten Kennzahl Befunde vorgelegt werden,²⁷⁾

- * die Auswertungsverfahren im allgemeinen nur der Logik des "Norm"-Vergleiches entsprechen, der menschliche Analytiker sich bei seiner Einzelfallbeurteilung aber vornehmlich auf Zeit- und Betriebsvergleiche stützt.

Diese Gründe erzwangen eine eigene Auswertung, um die **interne Validität** unserer Effizienzmessung zu sichern. Wir benötigen nämlich ein Verfahren zur Bestimmung der statistischen Relevanzmaße, das für alle Urteilsbereiche und Vergleichstechniken, die der Analytiker typischerweise vornimmt, in einheitlicher Weise angewendet werden kann. Außerdem sollen die Ziele und die Datensituation der statistischen Analyse der unseres Experiments möglichst gut entsprechen. Auf das Problem der externen Validität unserer Relevanzmaße werden wir bei der Diskussion der Ergebnisse eingehen.

Bei unserer Auswertung bestimmen wir zunächst für jeden Urteilsbereich eine **repräsentative Kennzahl**. Die Kennzahlen werden so gewählt, daß sie ein sachgerechtes Urteil für den jeweils zugeordneten Analysebereich ermöglichen und vergleichsweise leicht aus den angebotenen Informationen gebildet werden können.²⁸⁾

Als **Testfeld** verwenden wir eine Prüfstichprobe von 35 Unternehmen, die in der Wirtschaftspresse als Krisenfälle angesehen wurden und eine Kontrollstichprobe von 35 weiteren Unternehmen, die diesen in Rechtsform, Branche und Größe möglichst stark ähneln.²⁹⁾ Dieses Testfeld ist für unsere Zwecke besonders gut geeignet, weil auch die Fälle, die den Versuchspersonen vorgelegt werden, aus der Prüfstichprobe entnommen werden.

Die statistische Auswertung soll die Informationsmöglichkeiten der Versuchspersonen so gut wie möglich abbilden. Den Versuchspersonen werden bei jeder Analyse folgende Daten angeboten: die aktienrechtlichen Bilanzen und Gewinn- und Verlustrechnungen sowie daraus abgeleitete verdichtete Daten. Sie werden für das Analysejahr der jeweils zu beurteilenden Unternehmung, das Vorjahr dieser Unternehmung und für das Vorjahr einer Vergleichsunternehmung zur Verfügung gestellt. (Vgl. hierzu Kapitel 4.2.). Auf Basis dieses Angebotes können Norm-, Zeit- und Betriebsvergleiche vorgenommen werden.

Wir simulieren die experimentelle Angebotssituation, indem wir für jedes unserer 70 Unternehmen maschinell einen Zeitvergleich, einen Betriebsvergleich und einen Normvergleich durchführen. Als Analysejahr wählen wir dabei das Jahr der Publikation einer Unternehmenskrise. Im einzelnen werden folgende Meßwerte gebildet:

1. Beim **Zeitvergleich** wird bei jedem Analysefall die Veränderung **seines** Kennzahlenwertes zu seinem Vorjahreswert bestimmt. Wir simulieren diesen Vergleich dadurch, daß wir die **Differenz** zwischen dem Kennzahlenwert des Analysejahres und dem des Vorjahres bestimmen. Dies hat gegenüber einer **Steigerungsrate** den Vorteil, daß die statistische Analyse nicht durch Ausreißer verzerrt wird, wenn die Basisgröße sehr klein ist oder ein Vorzeichenwechsel stattfindet. Um eine derartige Berechnung zu ermöglichen, haben wir alle Kennzahlen als Quotienten ausgedrückt. Insbesondere werden alle Erfolgskennzahlen am Gesamtkapital oder an der Gesamtleistung relativiert, so daß man die Veränderung des Erfolges leicht an der Renditeveränderung in Prozentpunkten ablesen kann.
2. Beim **Betriebsvergleich** wird bei jedem Analysefall die Abweichung zu **seinem** Vergleichsfall bestimmt. Auch hierfür werden Differenzmaße gewählt. Abweichungen im Betriebs- und Zeitvergleich werden somit nach dem gleichen Rechenverfahren ermittelt.
3. Beim **Normvergleich** wird keine Differenzenbildung vorgenommen, sondern die Kennzahlenausprägung des Analysejahres wird den weiteren Auswertungen zugrundegelegt. Sie ist mit einer für alle Analysefälle identischen, statistisch zu bestimmenden Durchschnittsnorm zu vergleichen. Wir wählen dafür den Median, weil er am besten das typische, "mittelmäßig" ausgeprägte Unternehmen repräsentiert.

Für die Bestimmung der statistischen Relevanzmaße können unterschiedliche Verfahren herangezogen werden. Da die durchgeführten Tests sehr ähnliche Ergebnisse zeigen,³⁰⁾ verwenden wir als Relevanzmaß die **Klassifikationsgüte**. Sie ist definiert als der Anteil richtig klassifizierter Unternehmen und wird mit Blick auf die einzelnen Vergleichsarten folgendermaßen

bestimmt:

* Beim **Normvergleich** wird ein Unternehmen als "krisenbehaftet" ("krisenfrei") klassifiziert, wenn sein Kennzahlenwert im Analysejahr ungünstiger (günstiger) ausgeprägt ist als der Median aller Kennzahlenwerte des Analysejahres.

* Beim **Zeitvergleich** wird ein Unternehmen als "krisenbehaftet" ("krisenfrei") klassifiziert, wenn die Veränderung seines Kennzahlenwertes in Prozentpunkten vom Vorjahr zum Analysejahr ungünstiger (günstiger) ausgeprägt ist als der Median der entsprechenden Veränderung aller Kennzahlenwerte.

* Beim **Betriebsvergleich** wird ein Unternehmen als "krisenbehaftet" ("krisenfrei") klassifiziert, wenn sein Kennzahlenwert im Vorjahr ungünstiger (günstiger) ausgeprägt ist, als der seines ihm zugeordneten Vergleichsfalls.

Die Klassifikationsgüte ist ein sehr anschauliches Maß, das gegen Stichprobenveränderungen relativ robust ist, weil auf eine datenangepaßte Optimierung des Trennwerts verzichtet wird. Das Meßverfahren kann eine hohe interne Validität beanspruchen, weil man wohl davon ausgehen kann, daß sich die Erfahrung eines Analytikers in einer empirischen Norm niederschlägt, die durch den Median gut repräsentiert wird. Für die externe Validität des Maßes spricht, daß Urteile dann hoch gewichtet werden, wenn sie nicht nur bei dem analysierten Fall, sondern auch bei **anderen** Unternehmen zu einer zutreffenden Klassifikation führen.

Unsere nachfolgenden Befunde zeigen, wie die Klassifikationsgüten für die einzelnen Vergleichsarten und Urteilsbereiche ausgeprägt sind. (Tabelle 3.6. zeigt den aggregierten Befund. Die Detailergebnisse für die einzelnen Bereiche sind in den Tabellen 3.7. bis 3.9. dokumentiert).

1. Der **Betriebsvergleich** ist stets die **beste Vergleichstechnik**. Der **Zeitvergleich** liefert bei allen Urteilsblöcken das **schlechteste Ergebnis**. Der **Normvergleich** zeigt meist nur ein **mäßiges Ergebnis**.

2. Die Kennzahlen zur **Erfolgs- und Finanzlage** diskriminieren zuverlässiger als die Kennzahlen zur Vermögenslage. Es ergibt sich in diesem Punkt eine Parallele zur literarischen Relevanz der Urteilsbereiche.

Tabelle 3.6.: Klassifikationsgüte nach Vergleichsart und Analysefeld
(In Klammern: Standardabweichungen der Klassifikationsgüten)

Vergleichsart	Analysefeld			Signifikanz der Unterschiede*
	Erfolgs- lage	Vermögens- lage	Finanz- lage	
Normvergleich	58,1 (5,5)	55,1 (2,6)	57,3 (4,7)	0,094
Zeitvergleich	55,5 (4,1)	54,1 (3,2)	54,9 (4,1)	0,493
Betriebsver- gleich	61,3 (7,9)	56,9 (3,6)	63,0 (8,4)	0,042
Signifikanz der Unterschiede**	0,000	0,000	0,000	

* Rangvarianzanalyse für unverbundene Stichproben nach Kruskal-Wallis

** Rangvarianzanalyse für verbundene Stichproben nach Friedman

Tabelle 3.7.: Statistische Relevanz der Urteilsbereiche zur Erfolgslage

Nr.	Urteilsbereich	Norm- vergl.	Zeit- vergl.	Betr. vergl.	Alle Vergl.
1	Gesamterfolg	60,00	57,14	65,71	60,95
2	Steuern v. Einkommen, Ertrag, Vermögen	65,71	51,42	68,57	61,90
3	Ausschüttungen	65,71	54,28	71,42	63,80
4	Rücklagenentwicklung	62,85	57,14	71,42	63,80
5	Zinsaufwendungen	54,28	62,85	74,28	63,80
6	Ordentlicher Erfolg	71,43	60,00	71,43	67,62
7	Außerordentlicher Erfolg	65,71	60,00	71,42	65,71
8	Betriebserfolg	68,57	57,14	74,29	66,66
9	Finanzerfolg	62,85	57,14	51,42	57,13
10	Liquidationserfolg	62,85	57,14	62,85	60,94
11	Bewertungserfolg	65,71	54,28	71,42	63,80
12	Umsatzentwicklung	---	65,71	---	65,71
13	Materialeinsatz	57,14	51,42	62,85	57,13
14	Personal: gesamter Aufwand	51,42	60,00	51,42	54,28
15	Personal: Löhne und Gehälter	51,42	51,42	54,28	52,37
16	Personal: Soziale Abgaben	51,42	51,42	57,14	53,32
17	Personal: Altersv. u. Unterstützung	54,28	51,42	60,00	55,23
18	Abschreibg.: gesamter Aufwand	54,28	54,28	54,28	54,28
19	Abschreibg.: Sachanlagen	51,42	54,28	57,14	54,28
20	Abschreibg.: Verluste, Wertmindg.	60,00	51,42	57,14	56,18
21	Abschreibg.: Erträge aus Aufl. SoPo	58,57	55,71	51,42	55,23
22	sonstige Einsatzfaktoren	51,42	51,42	60,00	54,28
23	Erträge aus Gewinnbeteiligungen	57,14	52,85	68,57	59,52
24	Erträge aus Beteiligungen	57,14	50,00	57,14	54,76
25	Erträge aus and. Finanzanlagen	51,42	51,42	55,71	52,85
26	sonstige Zinsen u. ä. Erträge	57,14	54,28	54,28	55,23
27	sonstige Erträge ord. Anteil	54,28	57,14	51,42	54,28
28	Abschreibungen auf Finanzanlagen	54,28	52,85	52,85	53,32
29	Aufwendungen aus Verlustübernahme	57,14	58,57	55,71	57,14
30	Erträge aus Abgang Anlagevermögen	57,14	57,14	54,28	56,18
31	sonst. Ertr. außerord. Anteil	57,14	65,71	62,85	61,90
32	Einstellg. Sonderposten m. RLA	60,00	57,14	65,71	60,95
33	Erträge aus Zuschreibung AV	52,85	51,42	58,57	54,28
34	Herabsetzung Pauschal-WB Forderg.	52,85	51,42	52,85	52,37
35	Erträge aus Auflösung Rückstellg.	60,00	54,28	74,28	62,85

Tabelle 3.8.: Statistische Relevanz der Urteilsbereiche zur Vermögenslage

Nr.	Urteilsbereich	Norm- vergl.	Zeit- vergl.	Betr. vergl.	Alle Vergl.
36	Gesamtvermögen, Grobstruktur	54,28	57,14	51,42	54,28
37	Gesamtvermögen, Umschlag	54,28	51,42	57,14	54,28
38	Anlagevermögen, Grobstruktur	54,28	51,42	54,28	53,32
39	Anlagevermögen, Umschlag	54,28	57,14	60,00	57,14
41	Umlaufvermögen, Grobstruktur	54,28	57,14	57,14	56,18
42	Umlaufvermögen, Umschlag	51,42	51,42	51,42	51,42
44	Verbundaktiva	57,14	50,00	60,00	55,71
45	Grundstücke und Gebäude	51,42	54,28	54,28	53,32
46	Maschinen und maschinelle Anlagen	54,28	60,00	65,71	59,99
47	sonstige Sachanlagen, immaterielles AV	57,14	57,14	57,14	57,14
48	Beteiligungen	54,28	52,85	54,28	53,80
49	sonstige Finanzanlagen	54,28	54,28	52,85	53,80
50	Materiallager	54,28	51,42	57,14	54,28
51	Waren- und Erzeugnislager	57,14	60,00	58,57	58,57
52	Debitorenhaltung	54,28	51,42	54,28	53,32
53	monetäres Umlaufvermögen	54,28	51,42	60,00	55,23
54	sonstiges Umlaufvermögen	57,14	54,28	60,00	57,14
55	Investitionen	62,85	51,42	58,57	57,61

Tabelle 3.9.: Statistische Relevanz der Urteilsbereiche zur Finanzlage

Nr.	Urteilsbereich	Norm- vergl.	Zeit- vergl.	Betr. vergl.	Alle Vergl.
56	langfristige Deckungsrelationen	57,14	57,14	68,57	60,95
57	kurzfristige Deckungsrelationen	62,85	51,42	60,00	58,09
58	Cash-Flow ("dynamische Verschuldung")	68,57	62,85	80,00	70,47
59	Eigenkapitalausstattung	57,14	54,28	60,00	57,14
60	Fristigkeit der Finanzierung	54,28	62,85	54,28	57,13
61	Verbundpassiva	51,42	51,42	68,57	57,13
62	Grundkapital	54,28	51,42	54,28	53,32
63	Rücklagen ("Börsenkurs")	60,00	51,42	62,85	58,09
64	Sonderposten mit Rücklageanteil	60,00	54,28	67,14	60,47
65	Pensionsrückstellungen	54,28	51,42	51,42	52,37
66	andere Rückstellungen	51,42	51,42	57,14	53,32
67	Bankenverbindlichkeiten	60,00	54,28	77,14	63,80
68	Kreditorenhaltung	57,14	57,14	60,00	58,09
69	sonstige kfr. Verbindlichkeiten	54,28	57,14	60,00	57,14

Insgesamt gesehen fallen die durchschnittlichen Klassifikationsgüten recht niedrig aus. Dies kann auf drei Gründe zurückgeführt werden:

1. Es wird keine Insolvenz, sondern die Publikation einer **Unternehmenskrise** erklärt. Zwischen krisenfreien und krisenbetroffenen Unternehmen dürften wesentlich geringere Unterschiede bestehen als zwischen insolventen und solventen Unternehmen.
2. Es handelt sich um **einfache univariate** und nicht um multivariate Tests.
3. Als Trennwerte werden **keine stichprobenoptimierten Werte**, sondern Mediane oder Werte eines korrespondierenden Vergleichsfalls verwendet.
4. Die durchschnittliche Klassifikationsgüte vieler Urteilsbereiche ist ein **Mittelwert**, der auch Detailkennzahlen einbezieht, die zwar spezifische Krisenursachen des Einzelfalls anzeigen, aber über das gesamte Sample hinweg relativ schlecht diskriminieren.

Um die **externe Validität** der Relevanzmaße besser beurteilen zu können, wollen wir daher nun die zehn Urteilsbereiche betrachten, deren Kennzahlen im Durchschnitt aller drei Vergleichsarten die höchste Klassifikationsgüte aufweisen (vgl. Tabelle 3.10.).

In dieser Aufstellung nehmen die "dynamische Verschuldung" (Cash-Flow zu Fremdkapital) und die Kennzahlen zur Erfolgslage die ersten Plätze ein. Dies deckt sich mit unseren empirischen Ergebnissen zum gleichen Sample, bei denen wir einen längeren Zeitraum untersucht haben.³¹⁾ Es zeigt sich auch eine bemerkenswerte Übereinstimmung mit anderen empirischen Untersuchungen zur Insolvenzprognose, die univariate Befunde für mehrere Kennzahlen vorlegen und somit für einen Vergleich herangezogen werden können. Nach einer vergleichenden Auswertung von Krehl nehmen in jenen Untersuchungen Kennzahlen zur Erfolgslage und zum Innenfinanzierungsspielraum Spitzenplätze ein.³²⁾ In der vergleichenden Auswertung von 13 multivariaten britischen Studien von Taffler leistet in allen Diskriminanzformeln eine Erfolgs- oder Cash-Flow-Kennzahl einen signifikanten und praktisch relevanten Beitrag.³³⁾ Wir können somit die externe Validität unseres statistischen

Relevanzmaßes als gesichert ansehen, was die Rangplätze der skalierten Urteilsbereiche angeht. Bezüglich des Niveaus liegen unsere Klassifikationsgüten etwa in Höhe der Gütemaße, die bei Insolvenzprognosen für das fünfte bis dritte Jahr vor der Insolvenz erzielt werden.³⁴⁾ Dies läßt sich damit erklären, daß wir Unternehmenskrisen untersuchen, die im allgemeinen in einem beachtlichen Zeitraum vor einer Insolvenz stattfinden und auch keinesfalls in allen Fällen zu einer Insolvenz führen müssen.

Tabelle 3.10.: Aufstellung der zehn am besten diskriminierenden Bereiche

Rang	Urteilsbereich	Norm- ver- gleich	Zeit- ver- gleich	Betr.- ver- gleich	Alle Ver- gleiche
1	Dynamische Verschuldung	68.57	62.85	80.00	70.47
2	Ordentlicher Erfolg	71.43	60.00	71.43	67.62
3	Betriebserfolg	68.57	57.14	74.29	66.66
4	Außerordentlicher Erfolg	65.71	60.00	71.42	65.71
7	Bankenverbindlichkeiten	60.00	54.28	77.14	63.80
7	Ausschüttungen	65.71	54.28	71.42	63.80
7	Bewertungserfolg	65.71	54.28	71.42	63.80
7	Rücklagenentwicklung	62.85	57.14	71.42	63.80
7	Zinsaufwendungen	54.28	62.85	74.28	63.80
10	Ertr. a. Aufl. Rückstellg.	60.00	54.28	74.28	62.85

Neben diesen Gemeinsamkeiten mit vergleichbaren Untersuchungen sind die Besonderheiten unseres Befunds hervorzuheben. Sie betreffen die unterschiedlichen Klassifikationsgüten der einzelnen Vergleichsarten und die Abweichungen zwischen statistischem und literarischem Relevanzmaß.

Es fällt auf, daß die Klassifikationsgüten des **Betriebsvergleichs** stets am besten ausgeprägt sind. Diese Überlegenheit wird durch eine Rangvarianzanalyse über alle Urteilsbereiche auf einem sehr hohen Signifikanzniveau abge-

sichert (p kleiner als 0,0001). Handelt es sich hierbei um eine Bestätigung von Schmalenbachs geflügeltem Wort, daß beim Zeitvergleich nur "Schlendrian mit Schlendrian" verglichen würde, während beim Betriebsvergleich sich der Schlendrian nicht länger verbergen könne?³⁵⁾ Die Frage läßt sich mit unserem Material nicht abschließend beantworten, weil wir nur zwei aufeinanderfolgende Jahre verglichen haben, um der experimentellen Angebotssituation zu entsprechen.³⁶⁾ Bemerkenswert ist jedoch, daß sich der Betriebsvergleich nur auf die **Vorjahresdaten** bezieht und trotzdem deutlich besser abschneidet als der Norm- oder Zeitvergleich.

Die empirische Insolvenzforschung sollte hieraus Konsequenzen ziehen: Wenn man schon Problem- und Vergleichsfälle sorgfältig paart, dann sollte man auch paarweise Vergleiche durchführen und sich nicht darauf beschränken, alle Fälle an ein und demselben Normwert zu relativieren, wie es derzeit gängige Praxis ist.³⁷⁾

Für unsere Effizienzmessung ist bedeutsam, daß es sehr wohl auf die Vergleichstechnik ankommt, welche Treffsicherheit ein Analytiker erzielt. Es war notwendig, verschiedene Vergleichstechniken zu simulieren und unterschiedlich zu bewerten.

Bei den statistischen Relevanzmaßen nehmen Kennzahlen Spitzenplätze ein, die auch in der Literatur sehr häufig vorgeschlagen werden, z. B. die dynamische Verschuldung oder die betriebliche Rendite. Auf der anderen Seite wird die auf Rang sieben platzierte Kennzahl "Bankenverbindlichkeiten zu Gesamtkapital" nur sehr selten vorgeschlagen. Die beliebten Kennzahlen zur Grobstruktur des Vermögens, wie z. B. "Anlagevermögen zu Umlaufvermögen" treten in der Liste überhaupt nicht auf. Um die Übereinstimmungen und Abweichungen zwischen den verschiedenen Relevanzmaßen besser beurteilen zu können, haben wir deren Rangkorrelationen nach Spearman gebildet.

Die Rangkorrelationen sind zwar alle signifikant positiv ausgeprägt, aber sie nehmen recht geringe Werte an. Dies weist darauf hin, daß es zweckmäßig sein kann, **verschiedene** Treffsicherheitsmaße vorzuschlagen, weil die einzelnen Relevanzmaße empirisch unterschiedliche Anforderungen operationalisieren. Bemerkenswert ist auch, daß die Klassifikationsgüten der einzel-

nen Vergleichstechniken nur mäßig korrelieren. Dies deutet darauf hin, daß die Überlegenheit des Betriebsvergleichs nicht bei allen Urteilsbereichen gleich hoch ausgeprägt ist. (Vgl. hierzu die hohe Standardabweichung für die Klassifikationsgüte des Betriebsvergleichs in Tabelle 3.6.).³⁸⁾

Tabelle 3.11.: Zusammenhänge zwischen den Relevanzmaßen (Rangkorrelationen)

Statistische Relevanz	Statistische Relevanz			Literarische Relevanz
	Normvergleich	Zeitvergleich	Betriebsvergleich	
Normvergleich	1	0,31	0,57	0,26
Zeitvergleich		1	0,21	0,27
Betriebsvergleich			1	0,20

Die mäßige Korrelation der Vergleichstechniken eröffnet auf der anderen Seite die Möglichkeit, durch **Kombination** verschiedener Vergleichstechniken eine insgesamt zuverlässigere Aussage abzuleiten. So erreicht Krehl beim Betriebserfolg durch Kombination von Norm- und Zeitvergleich eine Klassifikationsgüte von 81,8 v. H., während bei uns diese beiden Vergleichsarten für sich genommen nur 68,5 bzw. 57,7 v. H. der Fälle korrekt klassifizieren.

Für unsere Effizienzmessung bedeutet die mäßige Korrelation der Klassifikationsgüten der verschiedenen Vergleichsarten, daß es sinnvoll ist, **pro Vergleichsart** ein eigenes Urteil in der Soll-Lösung vorzusehen, denn die Redundanzvermutung erweist sich als empirisch nicht stichhaltig.

3.2.3.3. Fallspezifische Relevanz

Die literarische und statistische Relevanzgewichtung liefern dem Analytiker eine Richtschnur, welche Analysebereiche vordringlich zu überprüfen sind und welche eher vernachlässigt werden dürfen. Sie kommen damit der begrenzten Kapazität des Analytikers entgegen. Aber: diese Regeln gehen von einer **durchschnittlichen** Relevanz der Urteilsbereiche aus. Sie berücksichtigen nicht, wie gravierend die Teilurteile im **Einzelfall** ausgeprägt sind. Für eine angemessene Würdigung des Einzelfalls kann aber die externe Ausprägung eines im Durchschnitt unwichtigen Tatbestands von ausschlaggebender Bedeutung sein. So besitzt die Veränderung der Positionen "Grundstücke und Gebäude" und "Maschinen und maschinelle Anlagen" im Durchschnitt nur eine mittlere Relevanz. Bei der Beurteilung des im Experiment zu analysierenden Falls Pegulan-Werke AG kommt aber der dramatischen Veränderung dieser Positionen im Zuge einer groß angelegten "Sale-and-lease-back"-Aktion eine zentrale Bedeutung zu. Wir müssen demnach trennen zwischen der a priori erwarteten Wichtigkeit der Analysebereiche und ihrer a posteriori festgestellten Relevanz für den Einzelfall.

Die Wichtigkeit dieses Qualitätsaspektes der Bilanzanalyse zeigt sich auch, wenn wir die Perspektive des Adressaten einnehmen. Auch seine **Aufnahme-** und **Verarbeitungskapazität** ist begrenzt. Die Bilanzanalyse bringt ihm nur dann eine Arbeitsentlastung, wenn der Analytiker eine **Filterung** vornimmt. So könnte man insbesondere daran denken, daß der Analytiker die a priori wichtig erscheinenden Urteilsbereiche **prüft**, aber nur die a posteriori im konkreten Einzelfall tatsächlich wichtigen Elemente **berichtet**.

Schließlich ist eine fallspezifische Würdigung der Kennzahlen erforderlich, um typische **Muster** zu erkennen und sachgerecht zu beurteilen.³⁹⁾ So zeigen sowohl krisenbehaftete als auch krisenfreie Unternehmen bei ungünstiger Branchenentwicklung rückläufige Betriebserfolge. Sie unterscheiden sich aber in der **Reaktion** auf diese Entwicklung: Krisenbehaftete Unternehmen mobilisieren in hohem Ausmaße Liquidations- und Bewertungserträge um Zins- und Steuerforderungen begleichen zu können, bei krisenfreien Unternehmen liegt das Erfolgsniveau im allgemeinen noch wesentlich höher, so daß sie den Erfolgseinbruch mit weniger offenkundigen bilanzpolitischen Maßnahmen

Wie kann man eine Norm für "fallspezifisch relevante Teilbestände" ableiten? Eine erste Möglichkeit könnte darin bestehen, Experten zu befragen und Gruppendiskussionen mit ihnen durchzuführen. Aber wer ist ein "Experte" und aufgrund welches Informationsstands und welcher Methoden kommen Experten zu ihrem Urteil?⁴⁰⁾ Wie kann man die Expertenurteile in Relevanzgewichte für unsere Urteilsbereiche transformieren? Eine zweite Möglichkeit könnte darin bestehen, festzustellen, welchen Beitrag eine repräsentative Kennzahl eines stärker desaggregierten Urteilsbereichs zu einer repräsentativen Kennzahl eines höher aggregierten, ihr unmittelbar hierarchisch übergeordneten Urteilsbereichs liefert. Dieses Verfahren würde jedoch voraussetzen, daß sich alle Ursache-Wirkungs-Beziehungen hierarchisch ordnen lassen und daß alle Analytiker genau diese Hierarchie anwenden sollten.

Wir haben uns daher für ein drittes Verfahren entschieden. Es lehnt sich an das von Schmidt vorgeschlagene Verfahren zur branchenbezogenen Lagebeurteilung an und umfaßt folgende Schritte:⁴¹⁾

1. Wie bei der statistischen Relevanzgewichtung wird pro Urteilsbereich eine repräsentative Kennzahl festgelegt (siehe oben).
2. Für die drei Vergleichsarten Norm-, Zeit- und Betriebsvergleich werden der Kennzahlenwert des Analysejahres, die Differenz zum Vorjahr und die Differenz zum Vergleichsfall als Ausgangswerte für jedes der 70 Unternehmen unseres Testfelds berechnet (siehe oben).
3. Für jede dieser Kennzahlenverteilungen werden Mittelwert und Standardabweichung ermittelt.
4. Das **fallspezifische Relevanzgewicht** eines Urteilsbereichs und einer Vergleichsart ergibt sich dann als Absolutbetrag des sogenannten "z-transformierten" Standardwerts:

$$FR_{ijk} = \frac{\text{abs}(K_{ijk} - K_{mit_{jk}})}{K_{stand_{jk}}}$$

mit:

FR	Fallspezifisches Relevanzgewicht
abs (...)	Absolutbetrag von (...)
K	Kennzahlenausprägung
Kmit	Mittlere Kennzahlenausprägung (arithmetischer Mittelwert der Kennzahlenausprägungen aller 70 Unternehmen)
Kstand	Standardabweichung der Kennzahlenausprägung
i	Index für die experimentellen Fälle (i = 1, ..., 4, K) (Der Index "K" steht für den Klausurfall).
j	Index für die Urteilsbereiche (j = 1, ..., 39, 41, 42, 44, ..., 69) (Anmerkung: Die Bereichsnummern 40, 43 und 70 bezeichnen Urteilsbereiche mit rein deskriptiven Aussagen und werden deshalb nicht berücksichtigt)
k	Index für die Vergleichsarten (k=1,2,3)

Dieses Verfahren ist sicherlich etwas pauschal.⁴¹⁾ Es hat aber den Vorteil, daß es allgemein anwendbar ist, einheitlich vorgeht, leicht nachvollzogen werden kann und auf objektiven empirischen Daten beruht. Der einzelne Urteilsbereich wird dann als "fallspezifisch relevant" angesehen, wenn er sich deutlich von der durchschnittlichen Kennzahlenausprägung eines Samples unterscheidet. Durch die vorgenommene Normierung der fallspezifischen Relevanzgewichte wird verhindert, daß die bereits bei den literarischen und statistischen Relevanzmaßen erfaßten Eigenschaften nochmals berücksichtigt werden. So werden in der Literatur hochaggregierte Kennzahlen mit größeren Werten besonders häufig vorgeschlagen. Mit der Division der Kennzahl durch ihre Standardabweichung wird dieser Größeneinfluß neutralisiert.

3.2.4. Die Prüfung der qualitativen Anforderungen an die Urteilsbildung

Unsere bisherigen Überlegungen bezogen sich auf ein **Soll-Konzept**. Es wurde dargelegt, welche Bereiche beurteilt werden sollen und wie man die Relevanz dieser Bereiche gewichten kann. Nun sind diesen Soll-Objekten die Ist-Urteile, die von den Analytikern getroffen werden, gegenüberzustellen. Ein derartiger Vergleich erfordert, daß die Ist-Urteile den gleichen Bereichen wie die Soll-Urteile zugeordnet werden und daß diese Messung in einer intersubjektiv nachvollziehbaren und zuverlässigen Weise erfolgt. (Vgl. hierzu Kapitel 3.1.).

Um dies zu gewährleisten, werden die von den Versuchspersonen abgegebenen bilanzanalytischen Urteile nach einem systematischen, am Institut für Betriebswirtschaftslehre der Christian Albrechts Universität zu Kiel eigens entwickelten inhaltsanalytischen Erhebungsverfahren ausgewertet.⁴²⁾

Als ein **Urteil** wird eine langschriftlich formulierte, verbale bilanzanalytische Aussage angesehen, die sich auf wenigstens eine numerische Jahresabschlußinformation oder eine aus diesen Daten errechnete Kennzahl stützt. Ein Urteil basiert immer auf einem explizit ausformulierten oder implizit erschließbaren **Vergleich** (Zeit-, Betriebs- oder Normvergleich).

Um die so definierten Ist-Urteile den Soll-Urteilen gegenüberzustellen, werden folgende Merkmale erhoben:

1. Der bilanzanalytische **Bereich**, auf den sich das Urteil bezieht (Merkmal "Analysebereich").
2. Die bilanzanalytischen Bereiche, mit denen der beurteilte Bereich **verknüpft** wird (Merkmale: "erste" und "zweite Verknüpfung").
3. Die **Positionen des Informationsangebotes**, die zur Ableitung des Urteiles herangezogen werden (Merkmal: "Positionsnummer").
4. Die **"Sachgerechtigkeit"** der Urteilsbildung.

5. Die Art der **Vergleichstechnik** (Norm-, Zeit- oder Betriebsvergleich).

Mit Hilfe der ersten drei Merkmale werden die Ist-Urteile den Bereichen der Soll-Urteile zugeordnet. Es handelt sich dabei um einen reinen Umschlüsselungsvorgang, der erforderlich ist, weil die Systematik der Soll-Urteilsbereiche von derjenigen abweicht, die bei der Erhebung der Ist-Urteile zugrundegelegt wurde.⁴³⁾ Maßgeblich hierfür war das Erfordernis, geeignete literarische und statistische Relevanzmaße zu finden. Die Systematik der Soll-Urteile sollte erstens den Differenzierungen der Literatur Rechnung tragen. Zweitens sollten die Urteilsbereiche so gebildet werden, daß die repräsentativen Kennzahlen, die bei der Bestimmung der statistischen Relevanzmaße verwendet werden, eine möglichst gleich hohe Diskriminationsleistung aufweisen wie andere Kennzahlen, die man diesem Bereich hätte zuordnen können. Diese Annahme wurde durch Alternativrechnungen mit anderen Kennzahlen empirisch geprüft. Auch das fünfte Merkmal dient zur Identifikation der Urteile, weil die statistischen und fallspezifischen Relevanzmaße von der Vergleichstechnik abhängen. Anzumerken ist, daß auch erhoben wird, welche Bewertung der Analytiker bei diesem Vergleich artikuliert. Dieses Merkmal wird bei der Operationalisierung der Beurteilungstendenz der Teilurteile im Kapitel 3.3. verwendet.

Von besonderer Bedeutung ist für uns das Merkmal "Sachgerechtigkeit". Es faßt die im Gliederungsabschnitt 3.2.1. vorgestellten qualitativen Anforderungen auf einer dreistufigen Skala mit den Ausprägungen "völlig", "teilweise" und "nicht sachgerecht" zusammen. Wir bilden eine zweistufige Skala mit den Ausprägungen "nicht" vs. "teilweise oder völlig sachgerecht", da die Anforderungen an die Sachgerechtigkeit sehr streng sind. Diese Zusammenfassung hat nur eine geringe Bedeutung, weil die mittlere Kategorie nur in seltenen Zweifelsfällen vergeben wird. Sie dient in erster Linie der Entlastung der Codierer, um solche Zweifelsfälle in einem zweiten Arbeitsgang eingehender zu würdigen.

Alle Ist-Urteile werden von einem sorgfältig geschulten Erheber nach schriftlich fixierten Regeln identifiziert und codiert. Ein zweiter Erheber, der sog. **Kollationierer**, prüft die Codierungen des Erhebers und korrigiert sie, falls erforderlich.⁴⁴⁾ Als Kollationierer werden zwei diplomierte

Projektmitarbeiter eingesetzt, die über besondere bilanzanalytische Erfahrungen verfügen. Sie garantieren die Einheitlichkeit der Anwendung der Codierungsregeln.

Unsere Prüfung der Reliabilität stellt daher nur eine Kontrollauswertung dar, um letztlich sicherzustellen, daß der Effizienzmessung verlässliche Daten zugrundegelegt werden. Zur Bestimmung der **Reliabilität** der Messung können drei Maßgrößen herangezogen werden:⁴⁵⁾

1. Die **Identifikationsrate**: Sie ist definiert als der Anteil der von Erheber und Kollationierer gemeinsam codierten Beurteilungen an der Vereinigungsmenge aller codierten Beurteilungen.
2. Die **Übereinstimmungsrate**: Sie ist definiert als der Anteil übereinstimmender Codierungen an der Menge der gemeinsam codierten Beurteilungen.
3. **Cohens Kappa**⁴⁶⁾: Dieses Maß ist schärfer als die Übereinstimmungsrate, weil es die beobachteten Übereinstimmungen an den bei Unabhängigkeit der Codierungen zufällig zu erwartenden Übereinstimmungen relativiert. Kappa ist definiert als:

$$K = \frac{\text{Beobachtete Übereinstimmungen} - \text{Erwartete Übereinstimmungen}}{\text{Gesamtzahl aller Codierungen} - \text{Erwartete Übereinstimmungen}}$$

Kappa drückt aus, um wieviel die beobachtete Übereinstimmung die aufgrund von Zufallseinflüssen durchschnittlich zu erwartende Übereinstimmung über- oder unterschreitet. Ein Wert von Null bedeutet, daß die Übereinstimmung gerade so hoch ist, wie man es aufgrund von zufälligen Einflüssen erwarten könnte. Ein Wert von Plus Eins bedeutet perfekte Übereinstimmung, ein Wert von Minus Eins völliger Dissenz.

Zur Bewertung unserer Ergebnisse sollen die Meßwerte vergleichbarer inhaltsanalytischer Studien vorangestellt werden. Wie Tabelle 3.12. zeigt, liegen die Übereinstimmungsraten im allgemeinen zwischen 0,70 und 0,90. Kappa streut um 0,65. Etwas schlechter fallen die Werte bei der Studie von

Angelmar und Stern aus. Dies dürfte daran liegen, daß soziale Interaktionen analysiert wurden und man bei der Codierung nicht auf ein fest definiertes Informationsangebot zurückgreifen konnte. Die selten berichtete Identifikationsrate beträgt in einem Fall 0,69, in einem anderen 0,91.

Für unsere Studie ergeben sich außer beim Merkmal "Sachgerechtigkeit" deutlich höhere Werte als in den anderen Untersuchungen⁴⁷⁾ (vgl. Tabelle 3.13.). Sie können auf folgende Ursachen zurückgeführt werden:

1. Die Kategorien werden präzise definiert und in einem umfangreichen Pretest ausgetestet und bereinigt.
2. Die Erheber werden gründlich geschult.
3. Der Kollationierer kann gegenüber einer unabhängigen Codierung Fehler vermeiden, da ihm die Unterlagen des Erhebers als Vorlage dienen.

Der dritte Einfluß wirft die Frage auf, ob die günstigen Meßwerte auf einem Vergleich von "Erheber-" und "Kollationiererschlendrian" beruhen. Dieser Einwand kann entkräftet werden: Eine frühere Auswertung von Krehl, die auf **unabhängigen** Codierungen mehrerer Erheber fußte, zeigte sehr ähnliche Reliabilitätswerte für die Codierung des Analysebereichs. (Die anderen Merkmale wurden in dem zugrundegelegten Pretest noch nicht codiert). Eine spätere Auswertung von Petersen zur Reliabilität der Reihenfolge der Codierungen ergab für unabhängige und abhängige Codierungen nahezu identische Ergebnisse.⁴⁸⁾

Der Einsatz von Kollationierern hat einen anderen Zweck als die Reliabilitätsprüfung, die in vielen anderen Studien vorgenommen wird. Es handelt sich bei uns nicht um eine methodische Prüfung an einer sehr kleinen Stichprobe, sondern um eine **vollständige** Prüfung und Korrektur aller erhobenen Ist-Urteile. Aus diesem Grund wählten wir das weitaus ökonomischere Verfahren mit einem Kollationierer statt einer **kompletten** unabhängigen Doppelcodierung aller Urteile.⁴⁹⁾ Die Reliabilität der Messung ist demnach **höher** als es die Meßwerte ausweisen, weil die erkannten Fehler **korrigiert** werden.

Tabelle 3.12.: Reliabilitätsmeßwerte vergleichbarer Studien

Quelle	Aufgabenstellung, Kategorienzahl	Identi- fika- tions- rate	Über- ein- stim- mungs- rate	Cohens Kappa
Angelmar/Stern (Bar- gaining 1978) S. 97f.	Verhandlungsprozesse, 12 Kategorien	0,69	0,66	0,62
Elstein/Shulman/Spraf- ka (Problem Solving 1978) S. 52	Medizinische Diagnose- prozesse, 8 Kategorien der Informationsbesch.	0,91	0,86	----
Elstein/Shulman/Spraf- ka (Problem Solving 1978) S. 155	Medizinische Diagnose- prozesse, Hypothesen- bildung, Info-Beschaffg.	----	0,80 bis 0,92 ^{a)}	----
Shields (Information Load 1978) S. 105	Kosten- und Leistungs- berichte, 20 Kategorien	----	0,71 bis 0,78	----
Bettman/Park (Ex- perience 1980) S. 239	Konsumgüterbeschaffung, 70 Kategorien	----	0,78	----
Russo/Johnson (Fami- liar Products 1980) S. 420	Konsumgüterbeschaffung, 42 Kategorien	----	0,83	----
Putz-Osterloh (Intel- ligenztestleistung 1981) S. 95	Intelligenztestaufgaben, 9 Kategorien	----	0,90 ^{a)}	----

Tabelle 3.12.: Reliabilitätsmeßwerte vergleichbarer Studien (Fortsetzung)

Quelle	Aufgabenstellung, Kategorienzahl	Identi- fika- tions- rate	Über- ein- stim- mungs- rate	Cohens Kappa
Shields/Birnberg/Frie- ze (Attributions 1981) S. 78	Leistungsbeurteilung von (fiktiven) Unterge- benen, 8 Kategorien	----	ca. 0,90 ^{b)}	----
Biehal/Chakravarti (Ex- periences 1982) S. 443	Konsumgüterbeschaffung, 70 Kategorien	----	0,90	----
Huber (Entscheiden 1982) S. 206	Auswahl von Ferienjobs, 8 Kategorien	----	----	0,75 ^{d)}
Biggs/Mock (Auditor Decision 1983) S. 239	Wirtschaftsprüfung, 14 Kategorien	----	0,75	0,66
Anderson (Process Tracing 1984) S. 163	Investitionsentschei- dung, 10 Kategorien	----	0,91	0,66
Biggs (Financial Ana- lysts 1984) S. 315	Aktienanalyse auf Basis von Jahresabschlüssen, 17 Kategorien	----	0,72 ^{a)}	0,67
Putz-Osterloh (Exper- ten 1986) S. 15	Unternehmensspiel, 21 Kategorien	----	0,77 ^{c)}	----

a) Auf Basis der Vereinigungsmenge aller Codierungen, sonst vermutlich auf Basis der Schnittmenge der von beiden Erhebern codierten Tatbestände

b) Intra-Coder Übereinstimmung nach 18 Monaten

c) Auf Basis der Schnittmenge der codierten Aussagen

d) Drei Codierer, verallgemeinertes Kappa nach Fleiss (Agreement 1971)

Tabelle 3.13.: Reliabilitätskennzahlen für die Erhebung der Ist-Urteile

Codierte Merkmale	Übereinstimmungsrate	Cohens Kappa
Objekt	0,81	0,80
1. Verknüpfung *	0,90	0,89
Positionsnummer **	0,97	0,96
Sachgerechtigkeit	0,86	0,62
Zeitvergleich	0,93	0,84
Betriebsvergleich	0,91	0,84
Normvergleich	0,91	0,84

Codierte Einheiten: 3029*** Identifikationsrate: 0,91

- * Die Reliabilitätsschätzung des Merkmals "1. Verknüpfung" beruht nur auf 114 gemeinsamen Codierungen. Eine "2. Verknüpfung" trat so selten auf, daß keine sinnvolle Reliabilitätsschätzung vorgenommen werden konnte.
- ** Die Reliabilitätsschätzung des Merkmals "detaillierende Positionsnummer" beruht auf 522 gemeinsamen Codierungen.
- *** Vereinigungsmenge aller codierten Beurteilungsaktivitäten der Experimentalserie Kiel 84, Analysefälle 1 bis 4

Die Sachgerechtigkeitsbeurteilung hängt nicht nur von der leicht zu prüfenden faktischen und logischen Wahrheit des Urteils, sondern auch von der **Angemessenheit der Kennzahlenbildung** ab. Eine solche "Angemessenheit" läßt sich nur mit Blick auf die **Meßziele**, die der Analytiker verfolgt, beurteilen. Da der diese nicht immer klar genug offenlegt, weil er sie nicht für erläuterungsbedürftig hält, oder weil sie ihm selbst nicht hinreichend klar sind, ist es nicht immer einfach, eine zuverlässige Einschätzung der "Angemessenheit der Kennzahlenbildung" vorzunehmen.

Um unsere Einschätzung zu objektivieren, haben wir uns daher entschlossen, drei Arten von Urteilsbereichen zu unterscheiden:⁵⁰⁾

1. Urteilsbereiche für sogenannte **"einfache Aussagen"**, bei denen der Analytiker kein spezifisches Analyseziel erkennen läßt, sondern lediglich die absoluten Ausprägungen von Angebotsinformationen oder deren absolute Abweichung vom Vergleichsfall oder Vorjahr stereotyp verbalisiert.
2. Urteilsbereiche für sogenannte **"qualifizierende Analysen"**. Sie zeichnen sich dadurch aus, daß der Analytiker durch ausdrückliche Formulierungen, die Art der Vorgehensweise oder die Art der gebildeten Kennzahlen eine spezifische Analyseabsicht zum Ausdruck bringt.
3. Urteilsbereiche für **"besondere Verknüpfungen"**. Sie zeichnen sich dadurch aus, daß eine kausale Ursachenforschung mit Hilfe mehrstelliger Relationen vorgenommen wird, wobei die verknüpften Informationen häufig aus unterschiedlichen Rechenwerken stammen. In diesem Fall ist die Angemessenheit der spezifischen Informationsverknüpfung zu bewerten.

Im Augenblick sollen uns nur die beiden ersten Arten von Urteilsbereichen interessieren. Die Urteilsbereiche der dritten Kategorie werden zur Operationalisierung der Effizienzdimension Qualität der Ursachenforschung verwendet. Auf sie werden wir im Kapitel 3.3. zurückkommen.

Die erste Art von Urteilsbereichen wurde geschaffen, um rein deskriptive Aussagen von qualifizierenden Urteilen abzugrenzen. Wenn beispielsweise formuliert wird: "Materialaufwand gestiegen, Personalaufwand gefallen, restlicher Aufwand unverändert", dann werden hierfür die Analysebereiche "423 Einfache Betrachtungen des Materialaufwands", "434 Einfache Betrachtungen des Personalaufwands" und "444 Einfache Betrachtungen des restlichen Aufwands" codiert. Die Aussage: "Für das schlechte Betriebsergebnis könnte die Ursache in Unwirtschaftlichkeiten des Materialbereiches gesucht werden, denn der Anteil des Materialaufwandes an der Gesamtleistung ist um 5 % gestiegen", wird demgegenüber dem Bereich "421 Qualifizierende Analysen zum Materialaufwand" zugeordnet.

In **beiden** Fällen hätte man jedoch eine "völlig" angemessene Kennzahlenbildung codiert. Im zweiten Fall, weil die Kennzahlenbildung dem Meßziel, ein Urteil über die Wirtschaftlichkeit des Materialeinsatzes zu geben, entspricht; im ersten Fall, weil kein Meßziel deutlich wird, an dem man die Angemessenheit hätte messen können. Diese unterschiedliche Codierung war vorgenommen worden, weil man zwischen zwei **unterschiedlichen Insuffizienzursachen** trennen wollte: einem Analytiker mit einem sehr geringen Anspruchsniveau, der sich von vornherein darauf beschränkt, Zahlen zu verbalisieren und einem anspruchsvollen Analytiker, der nicht in der Lage ist, für sein höhergestecktes Ziel angemessene Kennzahlen zu bilden.

Unsere Auswertungen zur Reliabilität der Messung zeigten jedoch, daß man die beiden Insuffizienzursachen empirisch nicht immer einwandfrei trennen konnte. Es ergibt sich eine deutlich höhere Reliabilität, wenn man beide Insuffizienzursachen zusammenfaßt. Tabelle 3.14. zeigt den Befund.

Durch die kombinierte Betrachtung steigt Cohens Kappa von 0,62 auf 0,71 an. Die Übereinstimmungsrate beträgt 0,87. Angesichts der schwierigen Codierungsproblematik sind beide Werte als gut anzusehen. Man kann die Teilmenge der anspruchsvollen und sachgerechten Urteile, die bei unserer Effizienzmessung verwendet wird, mit einer hohen Zuverlässigkeit bestimmen.

3.2.5. Die Meßwerte zur Treffsicherheit der Lagebeurteilung

Unsere Ausführungen zur Bestimmung der Urteilsbereiche, der Relevanzmaße und der Sachgerechtigkeit der Urteilsbildung haben deutlich gemacht, daß die "Treffsicherheit der Lagebeurteilung" ein vielschichtiges Konstrukt ist, das man in unterschiedlicher Weise messen kann. Wir wollen uns trotzdem auf **einen** Meßwert festlegen, der für die Beurteilung der Effizienzdimension "Treffsicherheit der Lagebeurteilung" maßgeblich ist, weil man sonst für eine Vielzahl unverbindlich nebeneinandergestellter Indikatoren Hypothesen ableiten und Befunde bewerten müßte.

Eine differenzierte Analyse der Ursachen hoher und niedriger Treffsicherheitsmeßwerte soll dadurch ermöglicht werden, daß der maßgebliche Hauptmeßwert systematisch nach verschiedenen Kriterien zerlegt wird, so daß sich ein hierarchisch aufgebautes System von Insuffizienz- und Effizienzmeßwerten ergibt. Damit soll die in Kapitel 3.1. aufgestellte Anforderung der "diagnostischen Eignung" erfüllt werden. Die weitere Analyse befaßt sich daher zunächst mit der Bildung des Hauptmeßwertes und wendet sich dann den Differenzierungen nach verschiedenen Kriterien zu.

Tabelle 3.14.: Kombination von Anspruchsniveau der Urteile und Angemessenheit der Kennzahlenbildung, beobachtete Codierungen (In Klammern: bei Unabhängigkeit erwartete Codierungen)

Kollationierer	Erheber			
	deskriptiv oder nicht angemessen	anspruchsvoll und angemessen	Summe	v. H.
deskriptiv oder nicht angemessen	848 (373,1)	233 (707,9)	1.081	37,1
anspruchsvoll und angemessen	157 (631,9)	1.674 (1.199,1)	1.831	62,9
Summe	1.005	1.907	2.912*	100,0
v. H.	34,5	65,5	100,0	

* Schnittmenge aller von Erheber und Kollationierer gemeinsam codierten Beurteilungsaktivitäten der Experimentalserie Kiel 84, Fälle 1 bis 4

3.2.5.1. Die Bestimmung des Hauptmeßwertes

Unsere Analyse der Relevanzmaße hat gezeigt, daß diese nur mäßig miteinander korrelieren. Es stellt sich daher die Frage, ob man überhaupt einen einheitlichen Treffsicherheitsmeßwert bilden sollte, oder man nicht besser zwischen "literarisch", "statistisch" und "fallspezifisch" bestimmten Varianten differenzieren sollte, weil die unterschiedlichen Gewichtungsverfahren andere Anforderungen an die Analytiker stellen und daher auch andere Fähigkeiten gemessen werden.

Auch wenn man von diesem Einwand absieht und alle drei Anforderungen in einem Meßwert bündelt, bleibt die Frage, wie die Relevanzmaße verknüpft werden sollen. Möglich wäre sowohl eine additive als auch eine multiplikative Verknüpfung der Relevanzgewichte.

* Bei einer **additiven Gewichtung** könnte man den Mittelwert aus "literarisch", "statistisch" und "fallspezifisch" gewichteter Treffsicherheit bilden.

* Die multiplikative Verknüpfung entspräche dem Vorgehen, das bei diskriminanzanalytischen Formeln zur Klassifikation von Unternehmen gewählt wird: Man multipliziert ein a priori Maß, den statistisch ermittelten Gewichtungssparameter einer Kennzahl, mit einem a posteriori Maß, der Kennzahlenausprägung eines zu klassifizierenden Unternehmens und summiert über die so gewonnenen Produkte. In unserem Fall können wir die statistischen und literarischen Relevanzgewichte als a priori Maße für Bereiche ansehen, die ein Analytiker prüfen sollte. Der fallspezifische z-Wert wäre das korrespondierende a posteriori Relevanzgewicht. Man könnte dieses mit dem Durchschnittswert der beiden a priori Maße multiplizieren.

In Abbildung 3.12. werden diese beiden Verknüpfungsvarianten der Relevanzgewichte formelmäßig definiert.

Abbildung 3.12.: Additive und multiplikative Berechnung der Treffsicherheit

$$TGA_{i1} = \frac{\sum_j \sum_k RLIT_j * A_{ijk1}}{3 * \sum_j \sum_k RLIT_j} + \frac{\sum_j \sum_k RSTA_{jk} * A_{ijk1}}{3 * \sum_j \sum_k RSTA_{jk}} + \frac{\sum_j \sum_k RFAL_{jk1} * A_{ijk1}}{3 * \sum_j \sum_k RFAL_{jk1}}$$

$$TGM_{i1} = \frac{\sum_j \sum_k \left(\frac{RLIT_j * A_{ijk1}}{2 * \sum_j RLIT_j} + \frac{RSTA_{jk} * A_{ijk1}}{2 * \sum_j \sum_k RSTA_{jk}} \right) * RFAL_{ijk1} * A_{ijk1}}{\sum_j \sum_k \left(\frac{RLIT_j}{2 * \sum_j RLIT_j} + \frac{RSTA_{jk}}{2 * \sum_j \sum_k RSTA_{jk}} \right) * RFAL_{ijk1}}$$

Hierbei bedeuten:

- i: Index für die Analytiker ($i = 1, \dots, 103$)
- j: Index für die Urteilsbereiche ($j = 1, \dots, 39, 41, 42, 44, \dots, 69$)
(Anmerkung: Die Bereichsnummern 40, 43 und 70 bezeichnen Urteilsbereiche mit rein deskriptiven Aussagen)
- k: Index für die Vergleichsarten ($k = 1, 2, 3$)
- l: Index für die experimentellen Fälle ($l = 1, \dots, 4, K$)
(Anmerkung: Der Indexwert "K" steht für den Klausurfall)
- A_{ijk1} : Bool'sche Variable, die den Wert Eins hat, falls der Analytiker i zum Urteilsbereich j und der Vergleichsart k beim Fall 1 ein qualifiziertes Urteil trifft. Andernfalls hat die Indikatorvariable den Wert Null.
- $RLIT_j$: Literarisches Relevanzgewicht für den Urteilsbereich j
- $RSTA_{jk}$: Statistisches Relevanzgewicht für den Urteilsbereich j und die Vergleichsart k
- $RFAL_{jk1}$: Fallspezifisches Relevanzgewicht für den Urteilsbereich j, die Vergleichsart k und den Fall 1

Um zu prüfen, ob diese Differenzierungen beachtlich sind, werden zwei Untersuchungen durchgeführt:

- a) Eine **Modellrechnung**, in der wir den Einfluß unterschiedlicher Gewichtungsverfahren auf der Basis genau festgelegter Prämissen studieren.
- b) Eine **konfirmatorische Faktorenanalyse**, in der überprüft wird, ob die unterschiedlichen Gewichtungsverfahren bei den von uns untersuchten Analytikern zu divergierenden Effizienzrangfolgen führen.

Zu a) Modellrechnungen zur Auswirkung unterschiedlicher Gewichtungsverfahren auf die Treffsicherheitswerte

Bei der Modellrechnung gehen wir von fünf fiktiven Analytikern aus, die sich im Leistungsniveau deutlich unterscheiden. Der jeweils bessere Analytiker fällt alle Urteile, die sein Vorgänger getroffen hat und beurteilt noch zusätzliche Aspekte. Zahl und Inhalt der Urteile werden so gewählt, daß die Urteilsmenge eine große Bandbreite abdeckt und die Analysefelder Erfolgs-, Vermögens- und Finanzlage auf allen Niveaus gleich gut repräsentiert sind. Um den Einfluß der Vergleichstechnik auf das Maßzahlenverhalten abschätzen zu können, werden alle Rechnungen für den reinen Zeitvergleich und den kombinierten Zeit- und Betriebsvergleich durchgeführt. Abbildung 3.13. zeigt unsere Vorgabe.

Mit dieser Modellrechnung sollen folgende Fragen beantwortet werden:

1. Welcher Meßwertverlauf ergibt sich für die einzelnen Gewichtungsverfahren in Abhängigkeit vom Leistungsniveau?
2. Wie wirken sich Falleinflüsse auf den Meßwertverlauf aus? Welcher Lernfortschritt wird angezeigt, wenn von Analysefall zu Analysefall eine Steigerung um ein Leistungsniveau stattfindet?

Abbildung 3.13.: Unterstellte qualifizierte Urteile für die Modellrechnung

Niveau	Erfolgslage	Vermögenslage	Finanzlage
1	Gesamterfolg, Ordentlicher Erfolg, Außerordentlicher Erfolg	Gesamtvermögen, Umschlag und Struktur	Deckungsrelation, lang- und kurzfristig, Eigenkapitalausstattung
2	Betriebserfolg, Finanzerfolg, Liquidationserfolg, Bewertungserfolg	Grobstruktur Anlage- und Umlaufvermögen	dynamische Verschuldung, Fristigkeit der Finanzierung
3	Umsatzentwicklung, Materialeinsatz, Personaleinsatz	Waren- und Erzeugnislager, Investitionstätigkeit	Rücklagen, Bankverbindlichkeiten, Lieferantenverbindlichkeiten
4	Ausschüttungen, Zinsen, Abschreibungen, sonstiger Aufwand	Materiallager, Debitorenhaltung	Pensionsrückstellungen, Andere Rückstellungen
5	Erträge aus Beteiligungen, Steuern, Erträge aus der Auflösung von Sonderposten mit Rücklageanteil, Einstellung in Sonderposten mit Rücklageanteil	Verbundaktiva, Beteiligungen	Verbundpassiva, Entwicklung steuerlich begünstigter Positionen mit Rücklagecharakter

Unser erster Befund zeigt das durchschnittliche Maßzahlenverhalten in Abhängigkeit vom Leistungsniveau. (Vgl. hierzu Abbildung 3.14. Die Durchschnittsbildung erfolgt über die Meßwerte der vier experimentellen Fälle und den Klausurfall).

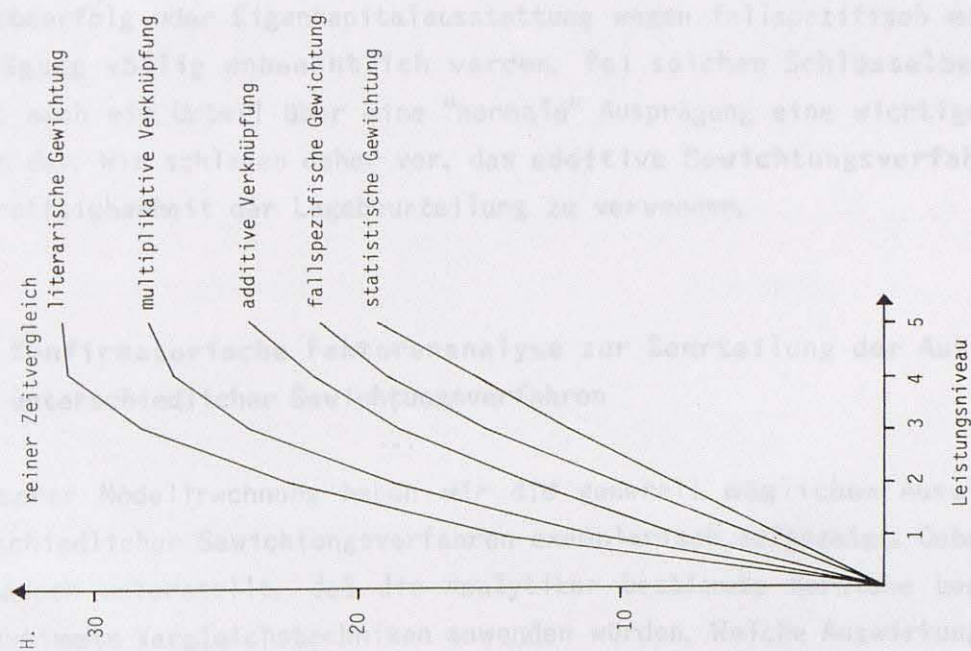
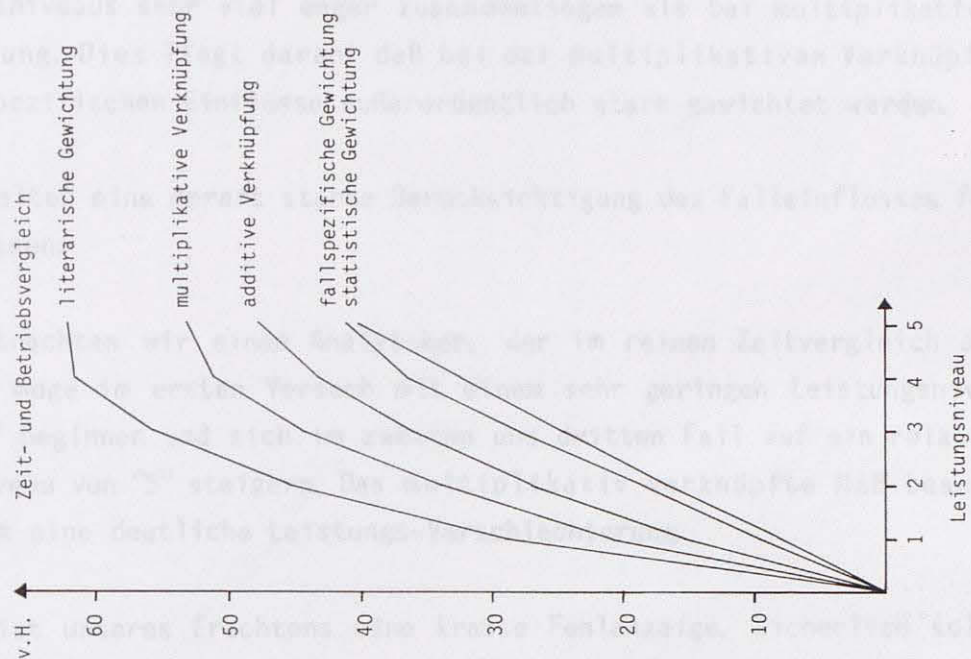
Für den reinen Zeitvergleich und den kombinierten Betriebs- und Zeitvergleich ergeben sich recht ähnliche Verläufe: Bei der literarischen Gewichtung steigt die Kurve zunächst sehr steil an, sie flacht danach deutlich ab. Bei der statistischen Gewichtung verläuft die Funktion praktisch linear. Die übrigen Verläufe liegen dazwischen. Die literarische Norm diskriminiert stark zwischen den unteren Leistungsniveaus. Sie zeigt nur sehr schwache Unterschiede zwischen den oberen Niveaus an.

Diese Unterschiede lassen damit begründen, daß in der bilanzanalytischen Literatur deutliche Präferenzen für bestimmte Urteilsbereiche geäußert werden, während sich die Klassifikationsgüten der Urteilsbereiche in wesentlich geringerem Ausmaß unterscheiden. Wie wir bereits gesehen haben, richten sich die Vorschläge der Literatur in erster Linie auf hochaggregierte Kennzahlen und in zweiter Linie auf die Neutralisierung von Störeinflüssen. Dadurch wird jedoch offenkundig nicht immer eine Verbesserung der Klassifikationsleistung erreicht, denn sonst hätten wir eine stärkere Korrelation zwischen beiden Relevanzmaßen feststellen müssen.

Auf der anderen Seite können auch Kennzahlen, die im Durchschnitt schlecht diskriminieren, anzeigen, welcher **Typ** einer krisenhaften Entwicklung vorliegt. Es wäre deshalb nicht unproblematisch, das statistische Relevanzmaß als alleiniges Gewichtungskriterium anzusehen.

Wir entscheiden uns daher, alle drei Relevanzmaße unserer Bewertung zugrunde zu legen. Die weitgehende Parallelität der Kurvenverläufe zeigt uns, daß wir damit ähnliche Anforderungen bündeln. Dies erhöht die Validität unseres Meßkonzeptes, weil wir uns nicht auf einen spezifischen Effizienzaspekt beschränken.⁵¹⁾

Abbildung 3.14.: Entwicklung der Treffsicherheitsmeßwerte bei unterschiedlichen Gewichtungsverfahren (erreichte Werte in v. H. der Maximalwerte)



Um zu beurteilen, ob für die Bündelung der Relevanzmaße eine additive oder eine multiplikative Verknüpfung besser geeignet ist, sehen wir uns nun den Befund zum Falleinfluß an (vgl. hierzu Abbildung 3.15.).

Man sieht, daß die Meßwerte bei additiver Verknüpfung bei gleichen Leistungsniveaus sehr viel enger zusammenliegen als bei multiplikativer Verknüpfung. Dies liegt daran, daß bei der multiplikativen Verknüpfung die fallspezifischen Einflüsse außerordentlich stark gewichtet werden.

Wir halten eine derart starke Berücksichtigung des Falleinflusses für unangemessen:

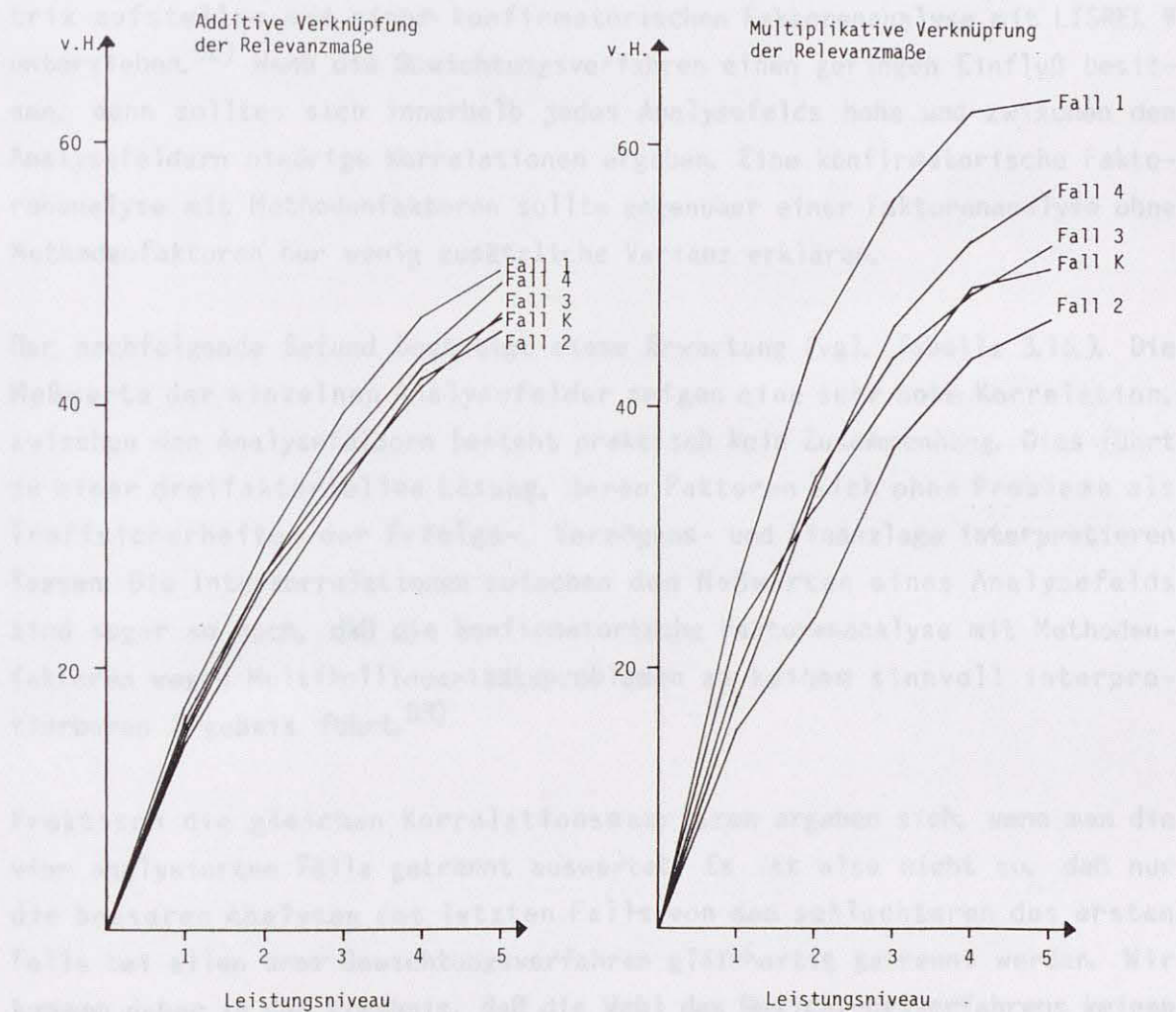
Betrachten wir einen Analytiker, der im reinen Zeitvergleich arbeitet. Er möge im ersten Versuch mit einem sehr geringen Leistungsniveau von "1" beginnen und sich im zweiten und dritten Fall auf ein relativ hohes Niveau von "5" steigern. Das multiplikativ verknüpfte Maß bescheinigt ihm eine deutliche Leistungs-**Verschlechterung**.

Dies ist unseres Erachtens eine krasse Fehlanzeige. Sicherlich sollte das Effizienzmaß fallspezifische Besonderheiten berücksichtigen und daher die Urteilsbereiche fallabhängig gewichten. Dies darf aber nicht dazu führen, daß **generell** für wichtig gehaltene Urteilsbereiche wie z. B. Gesamterfolg, Betriebserfolg oder Eigenkapitalausstattung wegen fallspezifisch mittlerer Ausprägung völlig unbeachtlich werden. Bei solchen Schlüsselbereichen stellt auch ein Urteil über eine "normale" Ausprägung eine wichtige Information dar. Wir schlagen daher vor, das **additive Gewichtungsverfahren** für die Treffsicherheit der Lagebeurteilung zu verwenden.

Zu b) Konfirmatorische Faktorenanalyse zur Beurteilung der Auswirkung unterschiedlicher Gewichtungsverfahren

In unserer Modellrechnung haben wir die generell **möglichen** Auswirkungen unterschiedlicher Gewichtungsverfahren exemplarisch aufgezeigt. Dabei haben wir jedoch unterstellt, daß die Analytiker bestimmte Bereiche beurteilen und bestimmte Vergleichstechniken anwenden würden. Welche Auswirkungen zei-

Abbildung 3.15.: Entwicklung der Treffsicherheitsmeßwerte bei unterschiedlichen Fällen und Gewichtungsverfahren (Zeit- und Betriebsvergleich, erreichte Werte in v. H. der Maximalwerte)



gen sich, wenn wir das **tatsächliche** Urteilsverhalten unserer Analytikerpopulation zugrundelegen?

Um diese Frage zu beantworten, haben wir für die drei Analysefelder Erfolgs-, Vermögens- und Finanzlage literarisch, statistisch und fallspezifisch gewichtete Treffsicherheitsmaße gebildet. Wir können damit in Anlehnung an Campbell und Fiske eine Multi-Objekt-Multi-Methoden Korrelationsmatrix aufstellen und einer konfirmatorischen Faktorenanalyse mit LISREL V unterziehen.⁵²⁾ Wenn die Gewichtungsverfahren einen geringen Einfluß besitzen, dann sollten sich innerhalb jedes Analysefelds hohe und zwischen den Analysefeldern niedrige Korrelationen ergeben. Eine konfirmatorische Faktorenanalyse mit Methodenfaktoren sollte gegenüber einer Faktorenanalyse ohne Methodenfaktoren nur wenig zusätzliche Varianz erklären.

Der nachfolgende Befund bestätigt diese Erwartung (vgl. Tabelle 3.15.). Die Meßwerte der einzelnen Analysefelder zeigen eine sehr hohe Korrelation, zwischen den Analysefeldern besteht praktisch kein Zusammenhang. Dies führt zu einer dreifaktoriellen Lösung, deren Faktoren sich ohne Probleme als Treffsicherheiten der Erfolgs-, Vermögens- und Finanzlage interpretieren lassen. Die Interkorrelationen zwischen den Meßwerten eines Analysefelds sind sogar so hoch, daß die konfirmatorische Faktorenanalyse mit Methodenfaktoren wegen Multikollinearitätsproblemen zu keinem sinnvoll interpretierbaren Ergebnis führt.⁵³⁾

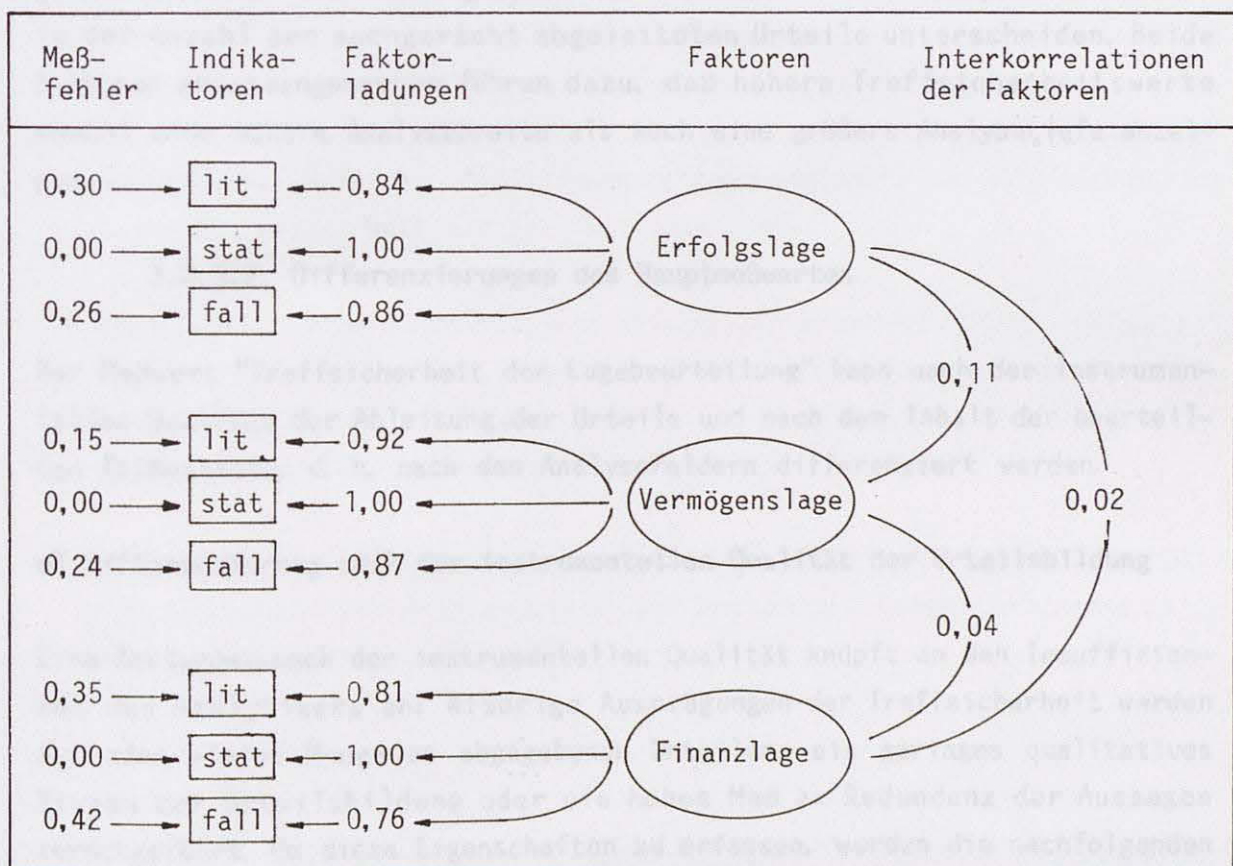
Praktisch die gleichen Korrelationsmatritzen ergeben sich, wenn man die vier analysierten Fälle getrennt auswertet. Es ist also nicht so, daß nur die besseren Analysen des letzten Falls von den schlechteren des ersten Falls bei allen drei Gewichtungsverfahren gleichartig getrennt werden. **Wir kommen daher zu dem Ergebnis, daß die Wahl des Gewichtungsverfahrens keinen Einfluß auf die intersubjektiven Effizienzunterschiede hat.**

Wie kommt das überraschende Ergebnis zustande? Wir sehen als Hauptgrund die große Homogenität des Urteilsverhaltens der Analytiker. Linsen-Modell-Studien mit amerikanischen und australischen Kreditmanagern zeigen einen hohen Konsens hinsichtlich der Gewichtung der angebotenen Kennzahlen.⁵⁴⁾ Weigel kann bei einem Sample von deutschen Kreditmanagern und unbegrenzten Nach-

Tabelle 3.15.: Korrelationsmatrix der Treffsicherheitsmeßwerte

Analysefelder		Erfolgslage			Vermögenslage			Finanzlage		
Gewichte		lit	stat	fall	lit	stat	fall	lit	stat	fall
Er- folgs- lage	lit	1,00								
	stat	0,84	1,00							
	fall	0,63	0,86	1,00						
Vermö- gens- lage	lit	0,04	0,02	0,00	1,00					
	stat	0,12	0,11	0,07	0,92	1,00				
	fall	0,10	0,10	0,10	0,81	0,87	1,00			
Fi- nanz- lage	lit	0,01	-0,09	-0,11	0,10	0,08	0,06	1,00		
	stat	0,09	0,02	-0,02	0,04	0,04	0,04	0,81	1,00	
	fall	0,04	-0,02	0,00	0,04	0,00	0,03	0,56	0,76	1,00

Abbildung 3.16.: Ergebnis der Konfirmatorischen Faktorenanalyse



Anmerkung: Die Meßfehler der statistischen Gewichtung wurden auf Null gesetzt, um sinnvolle Parameterschätzungen zu ermöglichen.

fragemöglichkeiten nachweisen, daß sich die Aussagen zu den Analysefeldern "Vermögen", "Kapital" und "Erfolg" recht gut auf einer Guttman-skala anordnen lassen.⁵⁵⁾ Diese Skaleneigenschaft bedeutet: Wenn ein Analytiker einen Tatbestand beurteilt, der nur von relativ wenigen Analytikern angesprochen wird, dann kann man mit großer Wahrscheinlichkeit davon ausgehen, daß er auch alle anderen Tatbestände beurteilt hat, die von sehr vielen Analytikern untersucht werden.

Hieraus folgt, daß ein umsichtiger Analytiker **unabhängig** vom Gewichtungungsverfahren besser eingestuft wird als ein weniger umsichtiger Analytiker. Wir vermuten, daß die Guttman-skalen-Eigenschaft auch in unserem Fall erfüllt ist, weil grundlegende Urteile, wie z. B. Aussagen zum Gesamterfolg praktisch von jedem Analytiker getroffen werden, während Begründungen, warum beispielsweise der Finanzerfolg stark gestiegen ist, von weitaus weniger Analytikern gegeben werden. Ein zweiter Faktor, der das beobachtete Ergebnis erklärt, kann darin gesehen werden, daß sich die Analytiker deutlich in der Anzahl der sachgerecht abgeleiteten Urteile unterscheiden. Beide Faktoren zusammengenommen führen dazu, daß höhere Treffsicherheitswerte sowohl eine höhere Analysebreite als auch eine größere Analysetiefe anzeigen.

3.2.5.2. Differenzierungen des Hauptmeßwertes

Der Meßwert "Treffsicherheit der Lagebeurteilung" kann nach der instrumentellen Qualität der Ableitung der Urteile und nach dem Inhalt der beurteilten Tatbestände, d. h. nach den Analysefeldern differenziert werden.

a) Differenzierung nach der instrumentellen Qualität der Urteilsbildung

Eine Zerlegung nach der instrumentellen Qualität knüpft an den Insuffizienzen des Analytikers an: Niedrige Ausprägungen der Treffsicherheit werden auf eine kleine Menge an abgegebenen Urteilen, ein geringes qualitatives Niveau der Urteilsbildung oder ein hohes Maß an Redundanz der Aussagen zurückgeführt. Um diese Eigenschaften zu erfassen, werden die nachfolgenden Effizienzindikatoren gebildet. Mit -1 multipliziert lassen sie sich als **Insuffizienzindikatoren** interpretieren.

Urteilsmenge = Anzahl aller abgegebenen Urteile

Qualitatives Niveau
der Urteilsbildung = $\frac{\text{Anzahl aller qualifizierten Urteile}}{\text{Anzahl aller abgegebenen Urteile}} * 100,0$

Redundanzfreiheit
der Teilurteile = $\left(1 - \frac{\text{Anzahl aller qualifizierten, nicht
wiederholt abgegebenen Urteile}}{\text{Anzahl aller qualifizierten Urteile}}\right) * 100,0$

Diese Meßwerte spiegeln die zunehmenden Anforderungen unseres Meßkonzeptes wider:

- * Die Urteilsmenge stellt eine unbereinigte Roheffizienz dar. Wenn ein Analytiker bereits bei diesem Maß einen sehr geringen Wert aufweist, dann braucht man erst gar nicht die Bewertungsprämissen, die wir der Sachgerechtigkeit der Urteilsbildung zugrundelegen, zu akzeptieren, um seine Treffsicherheit als gering einzustufen.
- * Niedrige Werte beim Qualitativen Niveau der Urteilsbildung deuten auf Flüchtigkeitsfehler oder mangelnden bilanzanalytischen Sachverstand hin. Pro Urteilsbereich und Vergleichsart ist in unserer Soll-Lösung höchstens ein qualifiziertes Urteil vorgesehen.
- * Sollte ein Analytiker nur wenige Urteilsbereiche sehr intensiv bearbeiten, dann kann man dies an niedrigen Werten des Indikators Redundanzfreiheit der Teilurteile ablesen.

b) Differenzierung nach dem Inhalt

Mit einer Zerlegung nach dem Inhalt kann geprüft werden, wie intensiv der Analytiker die Analysefelder Erfolgs-, Vermögens- und Finanzlage bearbeitet. Man kann feststellen, ob etwaige Lernfortschritte bei einzelnen Analysefeldern besonders groß sind, oder ob es bestimmte Typen von Analytikern gibt, die unterschiedliche inhaltliche Schwerpunkte setzen. Die Meßwerte werden in diesem Fall wie bei der Treffsicherheit der Gesamtlage gebildet. Die Summation erfolgt lediglich über eine kleinere Indexmenge der jeweils relevanten Urteilsbereiche.

Unter einer Zerlegung nach dem Inhalt versteht man die Aufgliederung der Urteilsbereiche in Teilbereiche, die nach dem Inhalt differenzierbar sind. Die Zerlegung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Zerlegung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Zerlegung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden.

Die Zerlegung nach dem Inhalt erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Zerlegung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Zerlegung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden.

3.3. Die Operationalisierung der Qualität der Urteilsanalyse

Die Operationalisierung der Qualität der Urteilsanalyse erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Operationalisierung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden.

Die Operationalisierung der Qualität der Urteilsanalyse erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Operationalisierung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden.

Die Operationalisierung der Qualität der Urteilsanalyse erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden. Die Operationalisierung erfolgt nach dem Inhalt der Urteilsbereiche, die in der Gesamtlage beurteilt werden.

3.3. OPERATIONALISIERUNG DER ANDEREN EFFIZIENZDIMENSIONEN UND DER BEURTEILUNGSTENDENZ

Mit der Effizienzdimension Treffsicherheit der Lagebeurteilung werden zwar die guten von den schlechten Analytikern auf einer fein abgestuften Skala zuverlässig und valide getrennt, es stellt sich jedoch die Frage, ob man zur Unterscheidung von guten und sehr guten Analytikern weitere Anforderungen heranziehen sollte. Die bisher operationalisierten Anforderungen erfassen das klassische Aufgabenverständnis der Bilanzanalyse, ein differenziertes Urteil über die Lage einer Unternehmung abzugeben. Unser Aufgabenverständnis ist weiter gefaßt: Die Bilanzanalyse soll nicht nur Entschlüsse begründen, sondern auch eine Initiativ- und Steuerungsfunktion für den nachfolgenden Entscheidungsprozeß übernehmen. Daher ist auch die Qualität der Ursachenanalyse und der entwickelten Fragen, die weitere Informationsaktivitäten unterstützen sollen, zu bewerten. Wir folgen damit den von Hauschildt vorgezeichneten Entwicklungslinien der Bilanzanalyse.¹⁾

Neben diesen beiden Effizienzdimensionen soll die Humaneffizienz operationalisiert werden. Schließlich soll gemessen werden, ob eine Analyse eine positive oder negative Beurteilungstendenz aufweist. Dabei handelt es sich nicht mehr um eine Effizienzgröße, sondern um eine beschreibende Maßgröße, die bei der Beziehungsanalyse im Kapitel 4.4. verwendet wird.

3.3.1. Die Operationalisierung der Qualität der Ursachenanalyse

Für komplexe Beurteilungsprobleme, wie z. B. Bilanzanalysen, werden nach Auffassung des Kognitionspsychologen Dörner und seiner Mitarbeiter unter anderem folgende geistigen Operationen benötigt:

- * **Trendanalysen:** Erkennen und Extrapolation von Verläufen in der Zeit. ("Wie entwickelt sich x?")²⁾
- * **Komponentenanalysen:** Zerlegung eines Ganzen in seine Teile. ("Woraus besteht x?")³⁾

* **Dependenzanalysen:** Hinterfragen kausaler Abhängigkeiten. ("Wovon hängt x in welchem Ausmaß und in welcher Form ab?")⁴⁾

Diese Anforderungen sind einem Betriebswirt wohl vertraut, gehört es doch zum Wesen der Bilanzanalyse, daß man Abweichungen von Vergangenheitswerten, von Ist-Werten anderer Vergleichsobjekte und von Normwerten feststellt und bewertet. Zur formalen Erklärung und präzisen Lokalisierung der Abweichungen werden Spitzenkennzahlen in nachgeordnete Zahlen zerlegt. Umgekehrt wird gefragt, welchen Einfluß bestimmte Komponenten auf das Ganze haben.⁵⁾

Die Qualität dieser Aktivitäten wurde bereits mit der Effizienzdimension Treffsicherheit der Lagebeurteilung erfaßt. Vordergründig gesehen handelt es sich hierbei um Operationen, die nur auf der Ebene der abgebildeten Zahlen stattfinden. Ein Bezug zu den zugrundeliegenden materiellen oder bilanzpolitischen Einflußfaktoren wird nicht zwingend vorausgesetzt. So kann die Komponentenanalyse als eine formale, Teil-Ganzes-Zerlegung angesehen werden, die lediglich auf definitorischen Zusammenhängen aufbaut. Auf der anderen Seite haben wir bei der Treffsicherheit der Lagebeurteilung bereits das Anspruchsniveau der bilanzanalytischen Absicht, die sachliche Angemessenheit der Kennzahlenbildung und die faktische Wahrheit der Urteilsprämissen bewertet. Es wurde also nicht nur eine logische Verknüpfungsleistung oder eine Rechenfertigkeit gemessen. Daher wird mit der Qualität der Ursachenanalyse kein völlig neues Konstrukt, sondern eine zusätzliche bilanzanalytische Leistung eingeführt, die auf einer treffsicheren Lagebeurteilung aufbaut.

* Verknüpfung der Entwicklung von Umsatze, Erzeugnisbestand, Forderungen

Mit der **Qualität der Ursachenanalyse** soll gemessen werden, in welchem Ausmaß der Bilanzanalytiker materielle oder bilanzpolitische Erklärungen aufstellt und überprüft. Bei der Überprüfung darf es sich nicht um eine reine Komponentenanalyse handeln, sondern um Dependenzanalysen, bei denen man mittels anspruchsvoller Verknüpfungen bilanzanalytischer Sachverhalte versucht, Hypothesen zu materiellen oder bilanzpolitischen Einflußfaktoren zu überprüfen. Die gemeinten Verknüpfungen zeichnen sich häufig dadurch aus, daß man Sachverhalte, die in unterschiedlichen Rechenwerken abgebildet werden, zueinander in Beziehung gesetzt werden und dabei mehrstellige Relationen bildet.

Zur Bedeutung solcher "Querbeziehungen" schreibt Richter:

"Eine nur punktuelle Betrachtung einzelner Posten (...) birgt die Gefahr von falschen Schlußfolgerungen und Fehlentscheidungen in sich. Bilanzierende, Prüfer und Bilanzleser sollten es sich deshalb zur Gewohnheit machen, stets die vielfältigen Beziehungen zwischen Bilanz und GuV und innerhalb beider Teile des Jahresabschlusses zu beachten und auch ggf. vorliegende Erläuterungen zum Jahresabschluß nicht zu vernachlässigen."⁶⁾

Was konkret unter den "besonderen Verknüpfungen" zu verstehen ist, die wir zur Messung der Qualität der Ursachenforschung heranziehen, soll anhand einiger Beispiele illustriert werden:⁷⁾

- * Verknüpfungen von Liquidationserfolgen mit außerordentlichen Vermögensverschiebungen und Veränderungen der Struktur der Fremdkapitalgeber, um die Hypothese zu testen, daß Vermögensteile zur Überbrückung von Finanzengpässen liquidiert werden.
- * Verknüpfungen von Anlageabgängen und Erträgen aus dem Abgang von Gegenständen des Anlagevermögens mit Veränderungen der sonstigen Aufwendungen um der Vermutung nachzugehen, ob eine "Sale-and-Lease-Back"-Transaktion durchgeführt wird.
- * Verknüpfung der Entwicklung von Umsatz, Erzeugnisbestand, Forderungen und Wertberichtigungen um zu prüfen, ob ein Absatzrückgang nicht nur erfolgsmindernd wirkt, sondern auch mit schlechter zahlenden Kunden verbunden ist.
- * Überprüfung der Rückstellungspolitik bei Erfolgsrückgang, um zu testen, ob Risiken in spätere Perioden verlagert werden.
- * Verknüpfung der Abschreibungen auf das Umlaufvermögen mit einem beobachteten Rückgang der Forderungen aus Lieferungen und Leistungen, um der Vermutung nachzugehen, daß ein ausgewiesener Debitorenrückgang nicht auf

kürzere Zahlungsfristen, sondern auf Forderungsausfälle zurückzuführen ist.

- * Verknüpfung der Pensionsrückstellungen mit der Entwicklung des Personalaufwandes, um zu ermitteln, ob die tatsächliche Steigerung noch höher ausgefallen wäre, wenn man nicht die Zuführung zu Pensionsrückstellungen unterlassen hätte.

Meßtechnisch wurden derartige Verknüpfungen über eigene Urteilsbereiche identifiziert.⁸⁾ Eine Restmenge von seltener aufgetretenen Verknüpfungen konnte über die Codierungsspalten erste und zweite "Verknüpfung" erfaßt werden.⁹⁾

Es fällt schwer, diese Verknüpfungen mit einem unterschiedlichen "Gewicht" zu belegen. Wir definieren daher schlicht folgenden Meßwert:

Qualität der Ursachenanalyse = Anzahl aller faktisch zutreffenden und sachgerecht abgeleiteten besonderen Verknüpfungen.

Auch wenn dieses Maß relativ grob ist, halten wir es für einen wichtigen Effizienzindikator. Nach den in Kapitel 3.1. belegten Befunden unterscheiden sich Anfänger und Experten der Bilanzanalyse vor allem dadurch, daß Experten komplexere Hypothesensysteme aufstellen und überprüfen. Die bereits referierten Ergebnisse der psychologischen Problemlöseforschung zeigen, daß dieses auch ein Merkmal erfolgreicher Problemlöser ist, die mit dynamischen, intransparenten Entscheidungen konfrontiert werden.

3.3.2. Die Operationalisierung der Qualität der entwickelten Fragen

Mit dem soeben vorgestellten Effizienzindikator "Qualität der Ursachenanalyse" wird nur ein Teil der Informationsaktivitäten zur Erforschung der Ursachen der beobachteten Lageentwicklung erfaßt. Bilanzen und Gewinn- und Verlustrechnungen sind nämlich wegen zahlreicher Beschränkungen ihres Datenmaterials nur bedingt in der Lage, Antworten auf vermutete Ursachen zu

geben. Dem gezielten Testen von Einflußhypothesen wird durch hoch aggregierte, unvollständige, vergangenheitsbezogene, gesetzlich normierte, spät bereitgestellte und bilanzpolitisch beeinflusste Daten eine recht enge Grenze gezogen, die auch durch geschickte Wahl von "besonderen Verknüpfungen" nur in begrenztem Umfang überschritten werden kann.¹⁰⁾ Die bilanzpolitische Ursachenforschung muß dann mit anderen Mitteln fortgesetzt werden. Sie muß versuchen, materiell begründete **Fragen**¹¹⁾ aufzuwerfen, die mit Hilfe anderer Informationsquellen beantwortet werden. Hierzu gehören beispielsweise der Geschäftsbericht, Presserklärungen des bilanzierenden Unternehmens, schriftliche und mündliche Informationen über das Unternehmen und nicht zuletzt die Auskünfte seiner Repräsentanten während der Kreditverhandlung.

In unseren Experimenten wird den Analytikern explizit aufgetragen, eine "Liste offener Fragen" zu formulieren (vgl. hierzu auch Kapitel 4.2.2.):

"Formulieren Sie eine Liste offener Fragen, die Sie aus den Bilanzmaterialien nicht beantwortet erhalten, damit ihr Chef in der nachfolgenden Kreditverhandlung mit der ... AG die Antworten direkt einholen kann."

In den Musterlösungen, die nach jeder Analyse vorgetragen wurden, wird ausführlich erläutert, welche Fragen jeweils nach einer Antwort drängen. Es sprechen daher sowohl theoretische Überlegungen als auch der Grundsatz der Angemessenheit der Effizienzbeurteilung dafür, die Qualität der Bilanzanalyse an den entwickelten Fragen zu messen.

Die **Frageleistung** soll die Leistung eines Analytikers messen, Fragen für eine Kreditverhandlung mit dem Bilanzersteller zu formulieren, die aus den angebotenen Bilanzen und Gewinn- und Verlustrechnungen nicht zufriedenstellend beantwortet werden können, für die Beurteilung der Erfolgs-, Vermögens- oder Finanzlage von großer Bedeutung sind und auf faktisch wahren Prämissen und logisch widerspruchsfreien Schlüssen beruhen.

Mit diesem Meßziel sind drei Mindestkriterien spezifiziert, die **unqualifizierte Fragen** ausschließen sollen:¹²⁾

1. Es darf sich nicht um rhetorische Fragen handeln, die leicht und gut aus den angebotenen Daten beantwortet werden können. So kann man beispielsweise die Frage: "Warum ist der Betriebserfolg so stark zurückgegangen?" relativ leicht dadurch beantworten, daß man den Betriebserfolg in seine ausgewiesenen **Komponenten** zerlegt und diese analysiert. Hingegen können die Fragen: "Warum ist der Umsatz so stark zurückgegangen? Handelt es sich um Absatzrückgänge, Preiseinbrüche oder Veränderungen des Produktionsprogramms?" nicht mehr aus den angebotenen Daten beantwortet werden.¹³⁾
2. Daß nur "wichtige" Tatbestände hinterfragt werden sollten, scheint eine plausible Anforderung an die Messung der Qualität der Frageleistung. Empirisch erwies sich dieses Kriterium als wenig beachtlich, weil praktisch nur Fragen zu Urteilsbereichen gestellt wurden, deren Relevanz nach den Maßstäben, die bei der Dimension Treffsicherheit entwickelt wurden, als hoch einzustufen ist. Die Detailanalyse der Restmenge zeigte überdies, daß durchaus beachtliche Detailinformationen in sehr präziser Form mit sehr spezifischen Begründungen verlangt wurden, während bei den Fragen zu den Urteilsbereichen mit hoher Relevanz teilweise recht global gefragt wurde: "Warum ist Position x so stark gestiegen?" oder: "Wieso ist Position y so viel höher als beim Vergleichsfall?" Daher haben wir das Relevanzkriterium **nicht** zur Abgrenzung qualifizierter Fragen herangezogen.
3. Wenn eine Informationsnachfrage auf falschen Prämissen beruht, also beispielsweise nach den Ursachen eines Anstiegs der Umsatzerlöse gefragt wird, obwohl die Erlöse de facto zurückgegangen sind, dann haben wir dies als eine unqualifizierte Frage angesehen. Eine Informationsaktivität, die auf solchen Fragen aufbaut, dient wohl mehr der Klärung von Irrtümern als der Gewinnung genuiner Erkenntnisse.

Aufgrund dieser Überlegungen definieren wir folgendes Effizienzmaß:

Allgemeine Frageleistung = Anzahl aller Fragen, die aus den angebotenen Informationen nicht beantwortet werden können und auf faktisch wahren Prämissen sowie logisch widerspruchsfreien Schlüssen beruhen.

Diese Menge von Fragen kann nach der Art der **Begründung** des Informationsbehrens differenziert werden. Die entwickelten Fragen verkörpern **Hypothesen**, die mit Hilfe des zur Verfügung stehenden Informationsmaterials nicht mehr getestet werden können. Man kann demnach die Qualität der Informationsnachfrage auch daran messen, wie präzise die Hypothese formuliert wird, die der Analytiker mit den gewünschten Informationen überprüfen will. Diese Sichtweise entspricht Wittes **"Erkenntnis"**-Erklärung, nach der auch eine **unversorgte** Informationsnachfrage zu einer Steigerung der Entscheidungsqualität führt, weil bereits die Formulierung einer Frage eine gewisse Problemstrukturierung voraussetzt.¹⁴⁾

Um diesen Qualitätsaspekt der Frageaktivität zu erfassen, wird jeweils festgestellt, ob es sich um eine pauschale oder um eine spezifisch begründete Frage handelt. Im zweiten Fall wird danach differenziert, ob materielle oder bilanzpolitische Einflußfaktoren als ursächlich angesehen werden.

Drei Beispiele sollen die Meßtechnik erläutern:

1. Pauschale Frage: "Warum ist der restliche Aufwand so stark gestiegen?"
2. Spezifische Frage, materieller Einflußfaktor wird vermutet: "Ist der restliche Aufwand so stark gestiegen, weil die Unternehmung einen Werbefeldzug durchgeführt hat, um dem Umsatzeinbruch zu begegnen?"
3. Spezifische Frage, bilanzpolitischer Einflußfaktor wird vermutet: "Sind die Erträge aus der Herabsetzung der Pauschalwertberichtigung zu Forderungen tatsächlich auf eine Verbesserung des kundenspezifischen Forderungsrisikos zurückzuführen oder handelt es sich um eine bilanzpolitische Maßnahme zur Verschönerung des Ergebnisausweises?"

Um diesen Effizienzaspekt zu erfassen, bilden wir einen zweiten Effizienzindikator, die "Spezifische Frageleistung":

Spezifische Frageleistung = Anzahl aller Fragen, die mit einer spezifischen Hypothese begründet werden, aus den angebotenen Informationen nicht beantwortet werden können und auf faktisch wahren Prämissen sowie logisch widerspruchsfreien Schlüssen beruhen.

Für dieses Effizienzmaß spricht nicht nur der dokumentierte Erkenntnisfortschritt des Analytikers, sondern auch die qualitativ bessere Informationsversorgung des Adressaten der Analyse, der nun eine Begründung erhält, warum er eine bestimmte Frage stellen sollte und sich damit besser auf das Gespräch vorbereiten kann.

3.3.3. Die Operationalisierung der Humaneffizienz

Die "Humaneffizienz" erfüllt in unserem Meßkonzept lediglich eine ergänzende und abrundende Funktion. Sie wird daher auch nur als subjektives Zufriedenheitsmaß vergleichsweise grob operationalisiert. Konkret werden den Versuchspersonen nach jeder Analyse folgende Fragen gestellt:

Zufriedenheit mit dem Ergebnis der Analyse: "Ich bin mit meinem Analyseergebnis sehr zufrieden". Likert-Skala mit den Polen "(1) stimme völlig zu", "(7) stimme überhaupt nicht zu".

Zufriedenheit mit dem Verlauf der Analyse: "Ich bin mit dem Verlauf meiner letzten Analyse sehr zufrieden". Likert-Skala mit den Polen "(1) stimme völlig zu", "(7) stimme überhaupt nicht zu".

Da die beiden Skalen sehr hoch miteinander korrelieren ($r = 0,81$), werden sie in der empirischen Analyse zu einem Faktor zusammengefaßt. Der leichteren Interpretation halber wird diese Skala so gepolt, daß größere Skalenergebnisse eine höhere Zufriedenheit anzeigen.

3.3.4. Die Operationalisierung der Beurteilungstendenz

Die Angemessenheit eines Gesamturteils spielt in den weiter oben referierten Linsen-Modell-Studien eine zentrale Rolle. Für uns ist das Gesamturteil nicht "die" Effizienzgröße, weil wir die Qualität der Entscheidungsvorbereitung messen wollen. Gleichwohl ist es von empirischer Bedeutung zu wissen, welche **Zusammenhänge** zwischen unseren Effizienzmaßen und der Beurteilungstendenz bestehen.

Unsere Operationalisierung stützt sich auf zwei Maße:

a) die Beurteilungstendenz des Gesamturteils,

b) die Beurteilungstendenz der Teilurteile.

3.3.4.1. Beurteilungstendenz des Gesamturteils

Eine explizite Empfehlung zur Annahme oder Ablehnung des Kreditantrages wird nur in 55 v. H. der erhobenen 412 Bilanzanalysen ausgesprochen.

Dieser Befund stützt sich auf eine Inhaltsanalyse, bei der die Erheber die Beurteilungstendenz einer langschriftlichen Aussage über die Gewährung oder Verweigerung eines Kredites beurteilen.¹⁵⁾ Konkret ist folgende Frage zu beantworten:

"Bitte geben Sie an, in welchem Ausmaß der Bilanzanalytiker eine Kreditvergabe befürwortet oder ablehnt:

- ☐ Er befürwortet die Kreditvergabe uneingeschränkt.
- ☐ Er befürwortet die Kreditvergabe tendenziell.
- ☐ Er macht keine Aussage zur Kreditvergabe.
- ☐ Er lehnt die Kreditvergabe tendenziell ab.
- ☐ Er lehnt die Kreditvergabe ausdrücklich ab."

Die Reliabilität dieser Einschätzung kann als gesichert angesehen werden: Die Übereinstimmungsrate beträgt 0,75, Cohens gewichtetes Kappa 0,86.

Bei der Serie **Kiel 84** werden die Versuchspersonen nach jedem Versuch explizit aufgefordert, ein **Urteil über die Kreditvergabe** auf einer siebenstufigen Likertskala abzugeben. Das entsprechende Item lautet:

"Ich befürworte eine Kreditvergabe" (1: "stimme völlig zu", 7: stimme überhaupt nicht zu")

Dies gibt uns die Möglichkeit zu prüfen, ob unsere Erheberbeurteilung verfahrensabhängig ist. Da die beiden Skalen unterschiedlich breit sind, können wir keine Übereinstimmungsrate und auch kein Kappa berechnen. Wir wählen stattdessen Kendalls Tau als Übereinstimmungsmaß für die beiden Ordinalskalen. Es beträgt 0,60 und bestätigt damit die **Konvergenzvalidität** unserer Messung.

Auch bei dieser erzwungenen Form einer Entscheidungsempfehlung sind die mittleren Kategorien der Skala sehr stark besetzt. Wir führen dies darauf zurück, daß die Ratingfrage offenkundig nicht dem subjektiven Aufgabenverständnis der Analytiker entspricht, bereits aufgrund einer Bilanzanalyse ein definitives Urteil über die Kreditvergabe zu fällen, denn die meisten Versuchspersonen machen in ihren Analysen die Entscheidung von der Beantwortung der "offenen Fragen" abhängig. Es ist daher nicht unproblematisch, die Beurteilungstendenz an einem solchermaßen ermittelten Gesamturteil zu messen.

3.3.4.2. Beurteilungstendenz der Teilurteile

Das Gesamturteil ist ein recht grob ausgeprägtes Maß, das die interindividuellen Unterschiede nur teilweise erfassen kann. Außerdem zeigen empirische Studien zum Urteilsverhalten von Experten, daß diese bei der Beurteilung einzelner Symptome sehr viel besser übereinstimmen als bei den Gesamturteilen, weil sie die Gewichtung der Symptome unterschiedlich vornehmen.¹⁶⁾ Daher ist es zweckmäßig, auch die Beurteilungstendenz der Teilur-

teile zu operationalisieren. Bei jeder Beurteilungsaktivität wird inhaltsanalytisch festgestellt, ob eine explizite Wertung vorgenommen wird und welche Richtung sie aufweist. Wir können daher folgende Maßgröße bilden:

$$\text{Beurteilungstendenz der Teilurteile} = \frac{\text{Negative} - \text{Positive Wertungen}}{\text{Negative} + \text{Positive} + \text{Neutrale Wertungen}} * 100,0$$

Unsere Auswertungen des Urteilsverhaltens sollen das Ausgangsmaterial beschreiben, auf das sich die weiteren Analysen stützen. Sie sollen darstellen, welche Bereiche beurteilt werden und wie sich das Urteilsverhalten bei den vier Fällen verhält. Die unter experimentellen Bedingungen ermittelten Urteilsverhaltensdaten werden im nächsten Kapitel (4.2.) analysiert.

Das Ausgangsmaterial wird dem Urteilsverhalten während des Klausurfalls gegenübergestellt, das unter normalen Bedingungen analysiert wird. Auf diese Weise können Effekte der Versuchssituation festgestellt werden.

In der vierten experimentellen Fallstudie wird auch das Urteilsverhalten von Studenten und Wirtschaftsprüfern analysiert, die an der Beurteilung des Urteilsverhaltens teilgenommen haben.

2. Kriteriumsvalidierung des Maßkonzepts hinsichtlich der Fähigkeit, Lernfortschritte differenziert zu diagnostizieren

Die Kriteriumsvalidierung soll zeigen, ob die bei der Beschreibung des Ausgangsmaterials ermittelten Lernfortschritte auch durch die Maßwerte unseres Effizienzkonzepts ausgedrückt werden. Sie soll offenlegen, bei welchen Maßwerten sich diese Lernfortschritte widerspiegeln.

3.4. EMPIRISCHE ANALYSE UND VALIDIERUNG DES MESSKONZEPTE

3.4.1. Überblick

Die Auswertungen des vorliegenden Kapitels dienen folgenden Zwecken:

1. Beschreibung des Urteilsverhaltens und Analyse von Einflüssen der Versuchssituation und der Versuchspopulation

Unsere Auswertungen zum Urteilsverhalten sollen das **Ausgangsmaterial beschreiben**, auf das sich die weiteren Analysen stützen. Sie sollen darstellen, **welche** Bereiche beurteilt werden und wie sich das Urteilsverhalten über die vier Fälle **entwickelt**, die unter experimentellen Bedingungen zu analysieren sind. (Zur Beschreibung der experimentellen Versuchsbedingungen siehe Kapitel 4.2.).

Dieses Ausgangsmaterial wird dann dem Urteilsverhalten während des Klausurfalls gegenübergestellt, der unter normalen Bedingungen analysiert wird. Auf diese Weise können **Einflüsse der Versuchssituation** festgestellt werden.

Da der vierte experimentelle Fall auch bei einem Wirtschaftsprüferexamen verwendet wurde, ist zusätzlich eine **Gegenüberstellung des Urteilsverhaltens von Studenten und Wirtschaftspraktikern**¹⁾ möglich.

2. Kriteriumsvalidierung des Meßkonzeptes hinsichtlich der Fähigkeit, Lernfortschritte differenziert zu diagnostizieren

Die Kriteriumsvalidierung soll zeigen, ob die bei der Beschreibung des Ausgangsmaterials erkennbaren Lernfortschritte auch durch die Meßwerte unseres Effizienzkonzeptes angezeigt werden. Sie soll offenlegen, bei **welchen Meßwerten** sich diese Lernfortschritte niederschlagen.

3. Konsistenzanalyse interindividueller Leistungsunterschiede bei verschiedenen Beurteilungsproblemen

Die Konsistenzanalyse soll die Frage beantworten, ob eine Analyseleistung bei einer Stichprobe von Testfällen x_1 bis x_{n-1} eine zuverlässige Prognose auf die Leistung bei einem weiteren Fall x_n ermöglicht.

Vor dem Hintergrund inkonsistenter Befunde²⁾ ist diese Frage von besonderer Bedeutung: Wenn es keine konsistenten Leistungsunterschiede zwischen verschiedenen Personen über verschiedene Fälle hinweg gibt, stellen die einzelnen Beurteilungsprobleme offenkundig spezifische Anforderungen. Es kann dann nicht "den" Experten für "die" Bilanzanalyse geben. Eine Suche nach dem für "die" Bilanzanalyse optimalen Kognitiven Stil wäre schon im Ansatz verfehlt. Das Forschungsinteresse muß sich vielmehr zunächst darauf konzentrieren, die Falleinflüsse zu generalisierbaren Aufgabenkategorien zu verdichten, um auf dieser Basis Empfehlungen für eine Entsprechung von Fähigkeit und Anforderung auszusprechen.

Falls es jedoch konsistente Leistungsunterschiede gibt, liefert eine Durchschnittsbildung der Effizienzmeßwerte über mehrere Fälle hinweg eine zuverlässigere und validere Leistungsbeurteilung. Unsere Auswertungen sollen zeigen, ob und wie groß der Qualitätsgewinn einer derartigen Durchschnittsbildung wäre.

4. Konstruktvalidierung des Meßkonzeptes

Bei einer Konstruktvalidierung³⁾ des Meßkonzeptes sind drei Prüfschritte simultan zu vollziehen:

- a) Überprüfung, ob Indikatoren, die derselben Effizienzdimension zugeordnet wurden, signifikant und hinreichend hoch miteinander korrelieren (**Konvergenzvalidität**).

- b) Überprüfung, ob Indikatoren, die verschiedenen Effizienzdimensionen zugeordnet wurden, niedriger miteinander korrelieren als die derselben Dimension untereinander (**Diskriminanzvalidität**).
- c) Überprüfung, ob die verschiedenen Effizienzdimensionen untereinander und mit einem Außenkriterium theoretisch prognostizierte Beziehungen aufweisen (**Nomologische Validität**).

Die einzelnen Schritte unseres Auswertungsprogramms bauen aufeinander auf und sollen in ihrer Gesamtheit die Validität unseres Meßkonzeptes prüfen.

Anzumerken ist, daß wir bereits bei der Gewinnung der Relevanzmaßstäbe auf die **Validität der Bewertungsprämissen** geachtet haben. Diese Validitätsproblematik ist eine Besonderheit einer Effizienzmessung und für ihre Gültigkeit von ausschlaggebender Bedeutung. Da sie bei der Operationalisierung anderer Konstrukte nicht auftritt, gibt es für diese "normative Validität" noch keinen eigenen Begriff.

3.4.2. Beschreibung des Urteilsverhaltens und Analyse der Einflüsse von Versuchssituation und Versuchspopulation

3.4.2.1. Beschreibung des Ausgangsmaterials

Bei unserer ersten Auswertung bestimmen wir, zu welchen Bereichen wieviel Prozent der Analytiker wenigstens ein qualifiziertes Urteil im Norm-, Zeit- oder Betriebsvergleich abgeben. (Siehe Tabelle 3.16. bis Tabelle 3.18.; Auswertungsbasis sind alle 103 Analytiker der Experimentalserien Gießen 82, Kiel 82 und Kiel 84).

Beim **ersten Fall** werden nur wenige Bereiche qualifiziert beurteilt: Nur Gesamterfolg, Umsatzentwicklung und Eigenkapital werden von mehr als der Hälfte der Analytiker kritisch gewürdigt. Die am häufigsten beurteilten Bereiche decken sich weitgehend mit den populärsten Kennzahlenvorschlägen der bilanzanalytischen Literatur. Abweichungen ergeben sich beim Cash-Flow

Tabelle 3.16.: Entwicklung der qualifizierten Urteile zur Erfolgslage

Nr.	Urteilsbereich	Analysierte Fälle				
		1	2	3	4	K
1	Gesamterfolg	54,4	77,7	84,5	74,8	96,8
2	StEEV	0,0	4,9	12,6	6,8	11,3
3	Ausschüttungen	0,0	13,6	3,9	7,8	29,0
4	Entwicklung Rücklagen	3,9	2,9	0,0	0,0	0,0
5	Zinsen	1,9	5,8	7,8	12,6	8,1
6	Ordentlicher Erfolg	1,0	2,9	2,9	2,9	1,6
7	Außerordentlicher Erfolg	8,7	7,8	6,8	3,9	12,9
8	Betriebserfolg	46,6	81,6	92,2	93,2	100,0
9	Finanzerfolg	15,5	38,8	76,7	77,7	88,7
10	Liquidationserfolg	10,7	49,5	35,9	81,6	80,6
11	Bewertungserfolg	14,6	61,2	80,6	33,0	85,5
12	Umsatzentwicklung	54,4	80,6	84,5	94,2	96,8
13	Materialeinsatz	8,7	52,4	56,3	56,3	88,7
14	Personaleinsatz	10,7	32,0	43,7	55,3	88,7
15	Löhne und Gehälter	1,0	6,8	4,9	13,6	53,2
16	Soziale Abgaben	0,0	3,9	4,9	12,6	35,5
17	Altersvers. und Unterstützung	1,0	3,9	7,8	10,7	53,2
18	Abschreibungen, insgesamt	3,9	14,6	26,2	27,2	69,4
19	Abschrbg. SAV und immat. AV	1,0	1,0	3,9	1,9	21,0
20	Verl. und Wertmindg. AV oder UV	0,0	1,9	3,9	1,0	33,9
21	Erträge a. Aufl. SoPo m. RLA	0,0	0,0	0,0	0,0	9,7
22	Sonstige Aufwendungen	9,7	27,2	45,6	32,0	62,9
23	Erträge aus Gewinngemeinschaften	1,0	0,0	3,9	2,9	4,8
24	Erträge aus Beteiligungen	0,0	0,0	8,7	1,9	0,0
25	Erträge aus and. Finanzanlagen	0,0	0,0	1,0	0,0	0,0
26	Sonstige Zinsen u. ä. Erträge	1,0	0,0	1,9	0,0	3,2
27	Sonstige Erträge, ord. Anteil	0,0	1,0	0,0	1,9	22,6
28	Abschreibungen auf Finanzanlagen	0,0	1,0	0,0	0,0	0,0
29	Aufwendungen a. Verlustübernahme	0,0	0,0	0,0	14,6	0,0
30	Erträge a. Abgang Anlagevermögen	1,0	4,9	0,0	36,9	1,6
31	Sonst. Erträge außerord. Anteil	0,0	0,0	0,0	20,4	11,3
32	Einst. SoPo mit Rückl.-Anteil	0,0	0,0	0,0	0,0	0,0
33	Erträge aus Zuschreibungen AV	0,0	10,7	20,4	0,0	1,6
34	Erträge aus Herabsetzung PWB	0,0	0,0	2,9	0,0	16,1
35	Erträge aus Auflösg. Rückstellg.	0,0	0,0	1,0	1,9	17,7

Lesebeispiel: 54,4 v. H. der Analytiker haben bei der Analyse des ersten Falls eine qualifizierte Beurteilung des Gesamterfolges vorgenommen. Bei den Fällen 1 bis 4 beziehen sich die Prozentzahlen auf 103 Analysen, beim Klausurfall "K" liegen 62 Analysen zugrunde.

Tabelle 3.17.: Entwicklung der qualifizierten Urteile zur Vermögenslage

Nr.	Urteilsbereich	Analysierte Fälle				
		1	2	3	4	K
36	Gesamtvermögen, Grobstruktur	26,2	20,4	12,6	7,8	29,0
37	Gesamtvermögen, Umschlag	2,9	1,9	1,0	0,0	1,6
38	Anlagevermögen, Grobstruktur	1,9	0,0	1,0	1,9	0,0
39	Anlagevermögen, Umschlag	0,0	2,9	1,0	0,0	3,2
41	Umlaufvermögen, Grobstruktur	1,0	1,0	1,9	1,0	0,0
42	Umlaufvermögen, Umschlag	1,0	1,0	0,0	0,0	0,0
44	Verbundaktiva	0,0	0,0	0,0	0,0	0,0
45	Grundstücke und Gebäude	1,9	0,0	0,0	0,0	1,6
46	Maschinen, Betr./Gesch.Ausstg.	1,0	0,0	0,0	1,0	0,0
47	Sonstiges Sachanlagevermögen	0,0	0,0	0,0	0,0	0,0
48	Beteiligungen	0,0	0,0	0,0	2,9	0,0
49	Sonstiges Finanzanlagevermögen	0,0	0,0	0,0	0,0	0,0
50	Materiallager	1,9	3,9	8,7	1,0	8,1
51	Waren- und Erzeugnislager	15,5	6,8	48,5	21,4	33,9
52	Debitorenhaltung	2,9	13,6	30,1	27,2	43,5
53	Flüssige Mittel	1,0	0,0	0,0	1,0	3,2
54	Sonstiges Umlaufvermögen	0,0	0,0	0,0	0,0	0,0
55	Investitionstätigkeit	14,6	13,6	31,1	28,2	80,6

Lesebeispiel: 26,2 v. H. der Analytiker haben bei der Analyse des ersten Falls eine qualifizierte Beurteilung der Grobstruktur des Gesamtvermögens vorgenommen. Bei den Fällen 1 bis 4 beziehen sich die Prozentzahlen auf 103 Analysen, beim Klausurfall "K" liegen 62 Analysen zugrunde.

Tabelle 3.18.: Entwicklung der qualifizierten Urteile zur Finanzlage

Nr.	Urteilsbereich	Analysierte Fälle					K
		1	2	3	4		
56	Langfristige Deckungsrelation	34,0	30,1	23,3	14,6		33,9
57	Kurzfristige Deckungsrelation	36,9	20,4	16,5	9,7		12,9
58	Cash Flow zu Fremdkapital	1,9	2,9	2,9	1,0		6,5
59	Eigenkapital	61,2	45,6	62,1	34,0		61,3
60	Langfristiges Kapital	17,5	20,4	19,4	54,4		30,6
61	Verbundpassiva	0,0	0,0	5,8	21,4		1,6
62	Grundkapital	1,9	3,9	1,9	0,0		0,0
63	Rücklagen	15,5	15,5	13,6	7,8		16,1
64	Sonderposten mit Rücklageanteil	1,0	0,0	0,0	0,0		1,6
65	Pensionsrückstellungen	6,8	10,7	22,3	12,6		38,7
66	Andere Rückstellungen	3,9	10,7	28,2	17,5		11,3
67	Bankenverbindlichkeiten	11,7	9,7	17,5	9,7		24,2
68	Kreditorenhaltung	5,8	8,7	5,8	13,6		27,4
69	Sonstige Passiva	4,9	15,5	4,9	2,9		11,3

Lesebeispiel: 34,0 v. H. der Analytiker haben bei der Analyse des ersten Falls eine qualifizierte Beurteilung der langfristigen Deckungsrelation vorgenommen. Bei den Fällen 1 bis 4 beziehen sich die Prozentzahlen auf 103 Analysen, beim Klausurfall "K" liegen 62 Analysen zugrunde.

und bei der Umsatzentwicklung. Dies liegt daran, daß Urteile zum Cash-Flow meistens als Aussagen zum Gesamterfolg codiert wurden. Nur wenn spezifische finanzwirtschaftliche Interpretationen des Cash-Flow vorgenommen und angemessene Kennzahlen für dieses Meßziel gebildet wurden, erfolgte eine Zuordnung zu diesem Urteilsbereich.⁴⁾ Daß die Umsatzentwicklung in der bilanzanalytischen Literatur einen geringeren Stellenwert einnimmt, ist wohl darauf zurückzuführen, daß viele Autoren sich auf solche "Kennzahlen" beschränken, die besondere Rechenoperationen erfordern, statt sich am betriebswirtschaftlichen Gehalt einer "Kennzahl" zu orientieren.⁵⁾

Zwischen der ersten und zweiten Fallbearbeitung findet die erste **zentrale Schulungsveranstaltung** statt. Sie bewirkt eine deutliche Veränderung des Urteilsverhaltens: Die Bereiche der Erfolgslage werden wesentlich häufiger qualifiziert beurteilt, bei der Vermögens- und Finanzlage ergibt sich eine Abnahme. Die während des Vortrags demonstrierte **Erfolgsquellenzerlegung**⁶⁾ schlägt sich in einer erheblich gestiegenen Beachtung der vier Erfolgsarten Betriebs-, Finanz-, Liquidations- und Bewertungserfolg nieder. Außerdem werden die Komponenten des Betriebserfolges häufiger analysiert. Diese Tendenz setzt sich in den nachfolgenden Analysen fort, bedingt durch positive Erfahrungen im Umgang mit dem Analyseinstrument und weiteren Schulungsveranstaltungen.

Im **dritten und vierten Fall** werden der Gesamterfolg und seine Hauptquellen von der breiten Mehrheit qualifiziert beurteilt. Außerdem würdigt ein beachtlicher Teil der Analytiker spezifische Komponenten des Betriebserfolges und der neutralen Erfolge. Bei der Vermögenslage werden Waren- und Erzeugnislager, Debitorenhaltung und Investitionsfähigkeit häufiger beachtet. Bei der Finanzlage gewinnen die Fristen- und Gläubigerstruktur an Bedeutung. Stark rückläufige Tendenz weisen die in den Schulungsveranstaltungen kritisch gewürdigten Deckungsrelationen auf.

3.4.2.2. Vergleich des Urteilsverhaltens bei den experimentellen Fällen und beim Klausurfall

Unser Befund stützt sich auf 64 Analysen der Serien Gießen 82 und Kiel 82, da bei der Serie Kiel 84 keine Klausurfälle erhoben wurden.⁷⁾ Wir rechnen beim Klausurfall mit einer nochmaligen Steigerung des Leistungsniveaus, weil:

- * das zeitaufwendige Ziehen, Ordnen und (teilweise) Übertragen von Informationskärtchen entfällt,
- * zusätzlich zum bisherigen Informationsangebot Steigerungsraten der verdichteten Erfolgspositionen bzw. Anteilswerte der verdichteten Bilanzdaten präsentiert werden,
- * die Versuchspersonen sich einen zusätzlichen Tag auf die Klausur vorbereiten können und vermutlich für die Klausur eine besondere Leistungsmotivation besitzen.⁸⁾

Unser Befund bestätigt diese Erwartungen: Die Anteile der qualifiziert beurteilten Bereiche steigen deutlich an (vgl. nochmals Tabelle 3.16. bis 3.18). Dies gilt insbesondere für die Komponenten des Betriebserfolges, die fallspezifisch stark ausgeprägten Komponenten der neutralen Erfolge, sowie die Komponenten des Umlaufvermögens und des Eigen- und Fremdkapitals. Außerdem erleben die in der Literatur vorgeschlagenen Kennzahlen zur horizontalen und vertikalen Bilanzstruktur eine beachtliche Renaissance.

Stellt man eine Rangliste der Urteilsbereiche auf, die sich auf den Durchschnitt der drei Fälle mit den jeweils größten Anteilen von qualifizierten Urteilen zu einem Bereich stützt, dann zeigt sich eine klare Dominanz der Urteilsbereiche zur Erfolgsanalyse (vgl. Tabelle 3.19.). Das Eigenkapital wird zwar noch von der Mehrheit der Analytiker beurteilt, rangiert aber erst auf Platz 9. Weitere Urteilsbereiche zur Vermögens- und Finanzlage treten erst ab den Plätzen 11 und 13 folgende auf.

Tabelle 3.19.: Rangliste der am häufigsten beurteilten Bereiche

Rg.	Urteilsbereich	Analysierte Fälle					
		1	2	3	4	K	D
1	Betriebserfolg	46.6	81.6	92.2	93.2	100.0	95.1
2	Umsatzentwicklung	54.4	80.6	84.5	94.2	96.8	91.7
3	Gesamterfolg	54.4	77.7	84.5	74.8	96.8	86.2
4	Finanzerfolg	15.5	38.8	76.7	77.7	88.7	80.9
5	Bewertungserfolg	14.6	61.2	80.6	33.0	85.5	75.7
6	Liquidationserfolg	10.7	49.5	35.9	81.6	80.6	70.5
7	Materialaufwand	8.7	52.4	56.3	56.3	88.7	67.0
8	Personalaufwand	10.7	32.0	43.7	55.3	88.7	62.5
9	Eigenkapital	61.2	45.6	62.1	34.0	61.3	61.4
10	Sonstige Aufwendungen	9.7	27.2	45.6	32.0	62.9	46.8
11	Investitionstätigkeit	14.6	13.6	31.1	28.2	80.6	46.5
12	Abschreibungen, insgesamt	3.9	14.6	26.2	27.2	69.4	40.8
13	Langfristiges Kapital	17.5	20.4	19.4	54.4	30.6	35.0
14	Waren- und Erzeugnislager	15.5	6.8	48.5	21.4	33.9	34.5
15	Debitorenhaltung	2.9	13.6	30.1	27.2	43.5	33.5
16	Langfr. Deckungsrelationen	34.0	30.1	23.3	14.6	33.9	32.6
17	Gesamtvermögen, Grobstruktur	26.2	20.4	12.6	7.8	29.0	25.1
18	Kurzfr. Deckungsrelationen	36.9	20.4	16.5	9.7	12.9	24.5
19	Pensionsrückstellungen	6.8	10.7	22.3	12.6	38.7	24.5
20	Löhne und Gehälter	1.0	6.8	4.9	13.6	53.2	24.4

Lesebeispiel: 54,4 v. H. der Analytiker haben bei der Analyse des ersten Falles eine qualifizierte Beurteilung des Gesamterfolges vorgenommen. Bei den Fällen 1 bis 4 beziehen sich die Prozentzahlen auf 103 Analysen, beim Klausurfall "K" liegen 62 Analysen zugrunde. Die Prozentangaben in Spalte "D" geben den Mittelwert aus den drei höchsten Werte einer Zeile an.

3.4.2.3. Vergleich des Urteilsverhaltens von studentischen Analytikern und Wirtschaftspraktikern

Zur Beurteilung der Problematik, ob die untersuchten studentischen Analytiker geeignete "Stellvertreter" für Wirtschaftspraktiker sind, können wir auf zwei empirische Untersuchungen zurückgreifen: eine Studie von Weigel zum bilanzanalytischen Informationsverhalten von Kreditmanagern und eine eigene Auswertung von Examensklausuren angehender Wirtschaftsprüfer.

Weigel führt mit 41 Kreditmanagern prozeßbegleitende Interviews, während derer die Kreditmanager durch "**lautes Denken**" darlegen, wie sie die Analyse eines konkreten Falls durchführen.⁹⁾ Zur Vorbereitung der Interviews werden den Kreditmanagern die Bilanzen und Gewinn- und Verlustrechnungen der (nicht genannten) Firma Kollmar und Jourdan AG, Pforzheim für die Jahre 1974 und 1975 zugesandt. Sie werden gebeten, diese Jahresabschlüsse mit Hilfe des institutsspezifischen Formulars aufzubereiten und eine Kopie des ausgefüllten Formulars an Weigel zurückzusenden. In der ersten Phase des Interviews wird das Jahr 1975 durchgesprochen. In der zweiten Phase wird das institutsspezifisch aufbereitete Jahr 1976 analysiert. Dabei wird den Kreditmanagern allerdings nicht das vollständig ausgefüllte Formular übergeben, sondern sie werden gebeten, die Werte der einzelnen Positionen zu erfragen. Die von wenigstens 75 Prozent der Interviewpartner erbetenen Informationen faßt Weigel zu einem "Basis-Satz relevanter Jahresabschluß-Positionen" zusammen (vgl. Tabelle 3.20.).¹⁰⁾

Sein Ergebnis ist mit unserem Befund nur bedingt vergleichbar, weil es sich auf die **Informationsnachfrage** nach wichtigen **Einzelpositionen** der aufbereiteten Bilanzen und Gewinn- und Verlustrechnungen bezieht und nicht auf qualifizierte **Urteile**, die auf angemessen gebildeten **Kennzahlen** beruhen. Daher fehlen bei Weigel Aussagen über horizontale und vertikale Bilanzstrukturrelationen, Rentabilitätskennzahlen und den Cash-Flow. Solche Kennzahlen sind aber nach einer weiteren Auswertung von Weigel in vielen Aufbereitungsformularen der Kreditinstitute vorgesehen.¹¹⁾ Sie werden auch in anderen empirischen Untersuchungen wiederholt als Schlüsselinformationen belegt.¹²⁾

Tabelle 3.20.: "Basis-Satz" relevanter Jahresabschlußpositionen

(Quelle: Weigel, Informationsverhalten 1980, S. 219)

Rangplatz	Erbetene Position	v. H. der Manager
1,5	Eigenkapital	97,6
1,5	Umsatzerlöse oder Gesamtleistung	97,6
3	Personalaufwand	95,1
4	Materialaufwand	92,7
5	Bankenverbindlichkeiten	90,2
6	Forderungen aus Lieferungen und Leistungen	78,1
7	Verbindlichkeiten aus Lieferungen und Leistungen	75,6
8	Wechselverbindlichkeiten	75,6

Gleichwohl erlaubt Weigels Befund einige Mindestaussagen:

1. Den Komponenten des Betriebsergebnisses: Umsatzerlöse oder Gesamtleistung, Materialaufwand und Personalaufwand wird auch von Kreditmanagern große Beachtung geschenkt.
2. Die in der Praxis stark nachgefragte Position Eigenkapital wird immerhin von 60 v. H. der studentischen Analytiker qualifiziert beurteilt.
3. Mit zunehmenden Lernfortschritten wenden sich die studentischen Analytiker auch den von Praktikern stark beachteten Positionen des Umlaufvermögens und des kurzfristigen Fremdkapitals zu. Bei der Bewertung dieser Tendenzen ist zu berücksichtigen, daß diese Positionen, insbesondere die Bankenverbindlichkeiten, beim Fall Kollmar und Jourdan besonders hoch ausgeprägt sind.¹³⁾ Das Nachfrageverhalten der Kreditmanager wird zudem stark auf diese Positionen gelenkt, weil sie zunächst zwei Jahresabschlüsse analysieren und dann erst die Daten eines dritten Abschlusses

nachfragen können.

4. Von beiden Gruppen werden die Positionen des Anlagevermögens stark vernachlässigt.¹⁴⁾

Die nächste Gegenüberstellung beruht auf einer Auswertung von 10 Examen-sklausuren angehender Wirtschaftsprüfer, die die Pegulan-Werke AG, den vierten experimentellen Fall, analysieren. Ihnen werden nur die aktienrechtlichen Bilanzen und Gewinn- und Verlustrechnungen zur Verfügung gestellt. Tabelle 3.21. belegt, welche Bereiche am häufigsten qualifiziert beurteilt werden.

Auch dieser Befund zeigt bemerkenswerte Übereinstimmungen beim Gesamt-, Liquidations- und Betriebserfolg sowie bei dessen Komponenten. Statt der "Aufwendungen aus Verlustübernahme" beurteilen die studentischen Analytiker meist den dramatischen Rückgang des Finanzerfolges, weil sie auf ein verdichtetes Informationsangebot zurückgreifen können. Beachtlich sind allerdings die Abweichungen bei den Verbundpassiva. Hierzu ist anzumerken, daß die Unternehmung bei einem Eigenkapital von 75 Millionen ein Bürgschaftsobligo an verbundene Unternehmen in Höhe von 140 Millionen DM ausweist. Diese fallspezifisch wichtige "Unter-Strich-Position" wurde den studentischen Analytikern nicht angeboten.

Fazit: Das Lehrziel, eine differenzierte Erfolgsquellenanalyse zu vermitteln, wird erfüllt. Wer das angebotene Konzept übernimmt, erlernt damit nicht nur eine akademisch für wichtig gehaltene Fähigkeit, sondern eine Fertigkeit, die in der Wirtschaftspraxis benötigt und verwendet wird. Wir sehen daher die Analysen unserer studentischen Versuchspersonen durchaus als geeignet an, Effizienzhypothesen zu überprüfen. Unsere Experimentalserien bieten darüber hinaus den besonderen Vorteil, die Entfaltung und Entwicklung bilanzanalytischer Fähigkeiten detailliert erforschen zu können.

Anmerkung: Bei den Wirtschaftsprüfern wurden wegen der kleinen Fallzahl keine Prozentzahlen berechnet. Die Werte sind absolute Angaben, d. h. alle 10 angehenden Wirtschaftsprüfer haben sich qualifiziert geäußert und Gesamterfolg getroffen. Die Werte der studentischen Analytiker beruhen auf 100 Analysen. Beide Populationen analysierten den Fall "Pegulan-Werke AG".

Tabelle 3.21.: Gegenüberstellung der von angehenden Wirtschaftsprüfern
und studentischen Analytikern beurteilten Bereiche

Urteilsbereich	Angehende Wirtschafts- prüfer (in von 10 Fällen)	Studenten der Wirt- schaftswis- senschaft (in v. H.)
Gesamterfolg	10	74,8
Liquidationserfolg	10	81,6
Umsatzentwicklung	10	94,2
Aufwdg. aus Verlustübernahme	10	14,6
Materialeinsatz	9	56,3
Ertr. a. Abgang Anlagevermögen	9	36,9
Verbundpassiva	9	21,4
Betriebserfolg	8	93,2
Beteiligungen	8	2,9
Langfristiges Kapital	8	54,4
Personaleinsatz	7	55,3
Gesamtvermögen, Grobstruktur	7	7,8
Eigenkapital	7	34,0
Lieferantenverbindlichkeiten	7	13,6

Anmerkung: Bei den Wirtschaftsprüfern werden wegen der kleinen Fallzahl keine Prozentzahlen berechnet. Die Werte sind absolute Anzahlen, d. h. alle 10 angehenden Wirtschaftsprüfer haben ein qualifiziertes Urteil zum Gesamterfolg getroffen. Die Werte der studentischen Analytiker beziehen sich auf 103 Analysen. Beide Populationen analysieren den Fall "Pegulan-Werke AG".

3.4.3. Entwicklung der Effizienzmeßwerte über die einzelnen Fälle

3.4.3.1. Entwicklung der Meßwerte zur Treffsicherheit der Lagebeurteilung

Wie entwickeln sich unsere **Meßwerte** zur Analysequalität? Zeigen sie den beobachteten Lernfortschritt an und wo schlägt er sich nieder?

Tabelle 3.22. dokumentiert die Entwicklung der Meßwerte zur Treffsicherheit der Lagebeurteilung, differenziert nach Analysefeldern. Der Befund belegt eine deutliche Zunahme der Treffsicherheit der Gesamtlage, die auf einem starken Zuwachs der Treffsicherheit der Erfolgslage beruht. Vergleicht man den ersten und zweiten mit dem dritten und vierten Fall, dann ergibt sich auch bei der Vermögenslage ein Lernfortschritt, der, wie wir gesehen haben, auf einer stärkeren Beachtung der Positionen des Umlaufvermögens beruht. Nur bei der Finanzlage ist ein Rückgang der Treffsicherheitsmeßwerte festzustellen. Er ist vornehmlich auf die gesunkene Beachtung kurzfristiger Deckungsrelationen zurückzuführen, die in der Literatur zwar relativ häufig vorgeschlagen, aber auch besonders heftig kritisiert werden.¹⁵⁾

Tabelle 3.23. belegt, daß diese Leistungssteigerung nachhaltig ist. Beim Klausurfall steigen alle Meßwerte weiter an. Die signifikante¹⁶⁾ Verbesserung der Treffsicherheit der Gesamtlage ist auf die starke Steigerung bei der Beurteilung der Vermögenslage zurückzuführen. Außerdem ist ein gewisser Einfluß der Versuchsserie festzustellen: Die Analytiker der Serien Gießen 82 und Kiel 82 beginnen auf einem deutlich höheren Niveau als die bei diesem Befund nicht berücksichtigten Teilnehmer der Serie Kiel 84.¹⁷⁾ Dies liegt daran, daß die meisten Teilnehmer der Serie Kiel 84 noch keine Veranstaltung zum Jahresabschluß im Hauptstudium besucht hatten. Trotz des höheren Ausgangsniveaus ergibt sich auch bei den Teilnehmern der Serien Gießen 82 und Kiel 82 eine signifikante Leistungssteigerung während der experimentellen Fälle, und zwar bei allen Meßwerten der Treffsicherheit.

Die Unterschiede der Treffsicherheit der Gesamtlage sind signifikant ausgeprägt. Sie entsprechen im Verlauf unseren Erwartungen. Man kann den Befund daher als **Kriteriumsvalidierung** unserer Meßwerte ansehen.¹⁸⁾

Tabelle 3.22.: Entwicklung der Treffsicherheitsmeßwerte, differenziert nach Analysefeldern (ohne Klausurfall, 103 Analysen pro Fall)

Analysefeld	Analysierte Fälle				Signifikanz ^{c)}
	1	2	3	4	
Gesamtlage	7,1 ^{a)} (4,2) ^{b)}	7,7 (3,0)	10,8 (3,6)	10,3 (3,1)	0,000
Erfolgslage	8,1 (6,6)	12,7 (5,0)	17,4 (5,8)	17,4 (6,4)	0,000
Vermögenslage	2,8 (3,8)	2,5 (5,0)	4,0 (4,1)	3,1 (3,6)	0,001
Finanzlage	10,0 (8,0)	6,2 (5,8)	7,7 (5,2)	6,7 (5,1)	0,003

Tabelle 3.23.: Entwicklung der Treffsicherheitsmeßwerte, differenziert nach Analysefeldern (mit Klausurfall, 64 Analysen pro Fall)

Analysefeld	Analysierte Fälle					Signifikanz	
	1	2	3	4	K	1 - K ^{d)}	4 vs. K ^{e)}
Gesamtlage	8,4 ^{a)} (4,0) ^{b)}	8,3 (3,2)	11,0 (3,5)	10,3 (3,0)	11,9 (3,2)	0,000	0,009
Erfolgslage	9,3 (6,7)	13,1 (5,3)	17,8 (5,5)	16,9 (6,3)	17,8 (4,1)	0,000	0,342
Vermögenslage	3,4 (4,0)	3,5 (5,4)	4,9 (4,5)	3,7 (3,8)	7,2 (4,0)	0,000	0,000
Finanzlage	12,0 (8,1)	6,6 (6,4)	7,5 (4,9)	7,1 (5,3)	8,4 (5,5)	0,001	0,134

a) arithmetisches Mittel, b) Standardabweichung, c) Rangvarianzanalyse nach Friedman über die vier experimentellen Fälle, d) Rangvarianzanalysen nach Friedman über die vier experimentellen Fälle und den Klausurfall, e) Wilcoxon-Matched-Sample-Test, vierter Fall vs. Klausurfall; zweiseitige Tests

Worauf gründet sich der beobachtete Lernfortschritt: Werden mehr bilanzanalytische Sachverhalte beurteilt? Steigt das Qualitative Niveau der Urteilsbildung oder die Redundanzfreiheit? Die Befunde in Tabelle 3.24. und 3.25. beantworten diese Fragen.

Die Effizienzsteigerung ist auf die wachsende Urteilsmenge zurückzuführen. Beim Qualitativen Niveau ergeben sich signifikante Unterschiede, aber der Verlauf ist uneinheitlich ausgeprägt. Die Redundanzfreiheit zeigt eine steigende Tendenz, doch die Streuungen sind zu hoch, um diesen Effekt auf einem vertretbaren Signifikanzniveau absichern zu können.

Durch die vorgetragenen Musterlösungen werden die Analytiker offenbar angeregt, mehr Sachverhalte zu beurteilen und tiefer nach ihren Ursachen zu forschen. Die Schulungsimpulse stoßen auf fruchtbaren Boden, weil die Analytiker durch Einübungseffekte Kapazitäten freisetzen. Welche Möglichkeiten das erlernte Konzept bietet, sieht man an der starken Zunahme der Urteilsmenge beim Übergang vom vierten experimentellen Fall zum Klausurfall: Die Zahl der durchschnittlich beurteilten Bereiche steigt von 27,7 auf 37,3 an. Dies entspricht einer Steigerungsrate von 34,7 Prozent!

Andererseits erfüllen nur rund 60 v. H. der Urteile die qualitativen Anforderungen. Von diesen 60 v. H. sind 15 v. H. Wiederholungen. Die "Wiederholungen" sind nicht unbedingt negativ zu werten, weil man sie für Zusammenfassungen benötigt und es zweckmäßig sein kann, bei einem Urteilsbereich mehrere Kennzahlen zu bilden. Aber der recht beachtliche Prozentsatz unqualifizierter gebildeter Urteile zeigt doch die Grenzen des Instrumentes "zentrale Schulungsveranstaltung" auf. Hier ist ein individuelles, auf den einzelnen Analytiker zugeschnittenes Feedback erforderlich. Die von uns entwickelten Insuffizienzindikatoren könnten dazu verwendet werden.

Fazit: Die untersuchten Analytiker erleben eine deutliche Steigerung der Treffsicherheitsmeßwerte. Ihre Analysequalität könnte aber durch umsichtigeren Vergleichstechnik und sachlich angemessenere Kennzahlenbildung noch verbessert werden. Hierzu sind vermutlich individuell zugeschnittene Lehrmethoden und längere Übungszeiträume erforderlich.

Tabelle 3.24.: Entwicklung der Treffsicherheitsmeßwerte, differenziert nach qualitativen Anforderungen (ohne Klausurfall, 103 Analysen pro Fall)

Indikatoren	Analysierte Fälle				Signifikanz ^{c)}
	1	2	3	4	
Treffsicherheit der Gesamtlage	7,1 ^{a)} (4,2) ^{b)}	7,7 (3,0)	10,8 (3,6)	10,3 (3,1)	0,000
Urteilsmenge	16,2 (11,1)	22,3 (9,3)	27,2 (9,3)	27,7 (8,6)	0,000
Qualitatives Niveau	58,4 (23,8)	58,2 (17,7)	63,6 (17,8)	57,3 (15,6)	0,010
Redundanzfreiheit	82,9 (15,6)	84,2 (12,9)	84,4 (10,5)	87,6 (10,1)	0,148

Tabelle 3.25.: Entwicklung der Treffsicherheitsmeßwerte, differenziert nach qualitativen Anforderungen (mit Klausurfall, 64 Analysen pro Fall)

Indikatoren	Analysierte Fälle					Signifikanz	
	1	2	3	4	K	1 - K ^{d)}	4 vs. K ^{e)}
Treffsicherheit der Gesamtlage	8,4 ^{a)} (4,0) ^{b)}	8,3 (3,2)	11,0 (3,5)	10,3 (3,0)	11,9 (3,2)	0,000	0,009
Urteilsmenge	18,2 (12,6)	23,4 (9,9)	28,9 (9,4)	27,7 (8,2)	37,3 (10,2)	0,000	0,000
Qualitatives Niveau	64,5 (21,6)	59,3 (16,9)	63,3 (19,0)	58,9 (15,7)	63,3 (12,6)	0,117	0,067
Redundanz-freiheit	81,3 (17,3)	84,0 (11,5)	83,6 (10,2)	88,1 (9,4)	84,5 (9,1)	0,136	0,016

a) arithmetisches Mittel, b) Standardabweichung, c) Rangvarianzanalyse nach Friedman über die vier experimentellen Fälle, d) Rangvarianzanalysen nach Friedman über die vier experimentellen Fälle und den Klausurfall, e) Wilcoxon-Matched-Sample-Test, vierter Fall vs. Klausurfall; zweiseitige Tests

3.4.3.2. Entwicklung der anderen Effizienzdimensionen und der Beurteilungstendenz

Tabelle 3.26. belegt folgende Entwicklungen: Die Qualität der Ursachenanalyse sowie die Allgemeine und Spezifische Frageleistung zeigen eine starke und signifikante Zunahme. Sie erfolgt vor allem während des dritten und vierten Falls. Es scheint, als ob die Analytiker zuerst lernen, welche Sachverhalte zu beurteilen sind, um sich dann die Fähigkeit anzueignen, nach den Ursachen auffällig ausgeprägter Bereiche zu forschen und weitergehende Fragen zu stellen. Erhärtet wird diese Vermutung zur Steigerung der **Analysetiefe** durch die Tatsache, daß die Qualität der Ursachenanalyse beim Klausurfall fast doppelt so hoch ist wie beim letzten experimentellen Fall.¹⁹⁾

Auch die Meßwerte zur Zufriedenheit zeigen einen signifikant positiven Trend. Die niedrigeren Werte beim vierten experimentellen Fall sind darauf zurückzuführen, daß dieser Fall als besonders schwierig eingeschätzt wird,²⁰⁾ was der Sachlage entspricht.²¹⁾

Die Urteilstendenz zeigt bei allen drei Meßwerten denselben, zwischen den Fällen signifikant unterschiedlich ausgeprägten Verlauf:

* Fall 2 (Maschinenbau-Balcke AG, 1969 und 1970) wird vergleichsweise günstig eingeschätzt,

* die Fälle 1 (Kollmar & Jourdan AG, 1975 und 1976) und 3 (Gerresheimer Glashüttenwerke AG, 1976 und 1977) werden sehr ungünstig beurteilt und

* Fall 4 (Pegulan-Werke AG, 1974 und 1975) wird vergleichsweise ungünstig eingestuft.

Diese Beurteilung wird sowohl durch die Ausprägung der Kennzahlen zur Erfolgs- und Finanzlage gestützt, als auch durch die Tatsache, daß bei der Firma Balcke das Jahr der Publikation einer Unternehmenskrise erst zwei Jahre nach dem Analysejahr liegt, während in den übrigen Fällen das Krisen-

Tabelle 3.26.: Entwicklung der Meßwerte der anderen Effizienzdimensionen und der Beurteilungstendenz (103 Analysen pro Fall)

Indikatoren	Analysierte Fälle				Signifikanz ^{a)}
	1	2	3	4	
Qualität der Ursachenanalyse	0,76 (1,27)	0,98 (1,38)	1,48 (1,75)	1,66 (1,89)	0,000
Generelle Frageleistung	2,04 (2,25)	3,01 (2,60)	4,96 (3,79)	5,32 (3,83)	0,000
Spezifische Frageleistung	1,25 (1,60)	1,54 (2,08)	2,74 (3,03)	2,79 (2,64)	0,000
Zufriedenheit mit Verlauf der Analyse	2,64 (1,46)	3,00 (1,55)	3,65 (1,43)	3,32 (1,57)	0,000
Zufriedenheit mit Ergebnis der Analyse	2,50 (1,42)	2,79 (1,47)	3,50 (1,39)	3,24 (1,49)	0,000
Beurteilungstendenz Teilurteile ^{b)}	35,8 (30,6)	28,7 (26,9)	46,7 (19,0)	34,5 (22,5)	0,000
Beurteilungstendenz Gesamturteil ^{b)}	3,37 (0,98)	2,86 (1,11)	3,81 (0,98)	3,38 (0,95)	0,000
Kreditvergaberating ^{b)} (nur Kiel 84)	4,44 (1,58)	3,94 (1,76)	5,14 (1,74)	4,72 (1,78)	0,030

a) Rangvarianzanalysen nach Friedman über die vier experimentellen Fälle

b) Höhere Werte zeigen eine stärker negativ ausgeprägte Beurteilung an

jahr als Analysejahr gewählt wird. Tabelle 3.27. zeigt, daß die Maschinenbau-AG Balcke bei der Erfolgslage klar am besten abschneidet. Lediglich das geringe Eigenkapital stellt einen Risikofaktor dar. Demgegenüber ist bei den anderen experimentellen Fällen sowohl die betriebliche Rendite als auch das Verhältnis von Cash-Flow zu Fremdkapital sehr schlecht ausgeprägt. Bei der Pegulan-Werke AG fällt lediglich der Eigenkapitalanteil und die langfristige Deckungsrelation besser aus als bei den übrigen experimentellen Fällen. Daß der erste Analysefall nicht genauso schlecht beurteilt wird wie der dritte, darf wohl auf die mangelnde Erfahrung der studentischen Analytiker bei diesem ersten Fall zurückgeführt werden.²²⁾

Hätten wir unsere Effizienzmessung nur auf die Angemessenheit der Beurteilungstendenz gestützt, dann hätten unsere Analytiker eine gute Note erhalten, weil sie ein zutreffendes Urteil abgeben. Da die Beurteilungstendenz nur ein vergleichsweise grobes und von subjektiven Risikopräferenzen abhängiges Maß ist, sehen wir sie nicht als ein Effizienzmaß an. Sie kann lediglich für eine grobe Validierung des Effizienzkonzeptes herangezogen werden.

3.4.4. Auswertungen zur Konsistenz interindividueller Leistungsunterschiede bei verschiedenen Fällen

Bei den Auswertungen zur Konsistenz interindividueller Leistungsunterschiede werden die analysierten Fälle als **Leistungstests** angesehen. Ein naheliegendes Konsistenzmaß ist die **Produktmomentkorrelation** zwischen den Leistungswerten zweier Fallbearbeitungen. Ein häufig verwendetes Maß für die Leistungskonsistenz bei mehr als zwei Fällen ist **Cronbachs Alpha**.²³⁾ Zur Bewertung von Alpha schreibt Nunnally, auf dessen Lehrbuch sich viele Autoren beziehen:

"What a satisfactory level of reliability is depends on how a measure is used. In the early stages of research ... reliabilities of .60 or .50 will suffice. (...) For basic research, it can be argued that increasing reliabilities beyond .80 is often wasteful."²⁴⁾

Tabelle 3.27.: Ausprägungen ausgewählter Kennzahlen der vier experimentellen Fälle und des Klausurfalls

		Betriebserfolg			Finanzerfolg			Außerordentlicher Erfolg		
		Kennzahlen-Ausprägungen			Kennzahlen-Ausprägungen			Kennzahlen-Ausprägungen		
		nvgl	zvgl	bvgl	nvgl	zvgl	bvgl	nvgl	zvgl	bvgl
1	Kollmar	-25.18	-26.75	26.42	7.43	5.57	0.74	5.33	4.56	-15.23
2	Balcke	-.42	-3.28	-1.42	1.14	-.37	-.14	1.26	0.77	-.52
3	Gerresh.	-12.79	-9.81	-5.35	8.96	1.56	2.11	3.90	2.84	-.70
4	Pegulan	-10.42	-12.56	-6.19	-.71	-8.22	3.09	18.93	16.76	1.17
K	Phoenix	-1.64	-.66	-3.79	4.18	1.36	0.39	3.26	2.08	-5.74
	Mittelw.	-3.15	-1.73	0.00	3.80	0.12	0.00	3.88	1.40	0.00
	Std.Abw.	11.94	8.89	17.41	5.65	5.89	6.60	5.67	4.21	5.47
	Minimum	-45.22	-28.67	-72.57	28.44	-28.62	-26.85	-2.11	-5.92	-15.23
	Maximum	31.17	24.00	72.57	21.29	27.39	26.85	23.89	19.22	15.23

		Cash-Flow zu Fremdkapital			Lgfr. Deckungsrelation			Eigenkapitalanteil		
		Kennzahlen-Ausprägungen			Kennzahlen-Ausprägungen			Kennzahlen-Ausprägungen		
		nvgl	zvgl	bvgl	nvgl	zvgl	bvgl	nvgl	zvgl	bvgl
1	Kollmar	-21.07	-25.37	23.95	92.18	-115.05	175.34	17.16	-17.44	21.16
2	Balcke	2.22	-3.41	-4.66	103.64	-4.62	-22.10	16.29	-3.24	-19.63
3	Gerresh.	-.99	-12.66	-20.69	58.53	-11.67	-13.00	28.21	-6.29	17.69
4	Pegulan	-24.20	-30.13	-25.23	111.25	16.05	-35.95	35.71	0.72	-18.23
K	Phoenix	15.92	9.34	-8.31	148.40	26.81	48.58	34.57	3.16	-5.72
	Mittelw.	6.59	-2.49	0.00	77.58	-2.60	0.00	23.81	-1.84	0.00
	Std.Abw.	21.48	14.89	24.11	60.29	33.10	73.88	19.32	7.44	20.19
	Minimum	-57.18	-53.35	-86.71	-224.63	-128.63	-204.13	-83.41	-41.86	-73.56
	Maximum	102.45	42.41	86.71	219.65	160.10	204.13	56.83	15.94	73.56

Legende:

nvgl (Normvergleich): Ausprägung der Kennzahl im Analysejahr

zvgl (Zeitvergleich): Kennzahlenwert Analysejahr minus Kennzahlenwert Vorjahr

bvgl (Betriebsvergleich): Kennzahlenwert Vorjahr Analysefall minus Kennzahlenwert Vorjahr Vergleichsfall

1, 2, 3, 4: Laufende Nummer der experimentellen Fälle, in der Reihenfolge der Bearbeitung; K: Klausurfall

Die Statistiken beziehen sich auf die 35 Fallpaare unseres Testfelds, als Analysejahr wird das Jahr der Publikation der Unternehmenskrise gewählt. Vgl. hierzu auch die "Liste der Unternehmen, die zur empirischen Prüfung der Analyse-Designs herangezogen wurden" bei Hauschildt (Erfolgs- und Finanz-Analyse 1985).

Bei der Übernahme dieser Standards ist zu bedenken, daß Alpha mit wachsender Zahl von Items zunimmt und wir nur wenige Fälle als Testitems verwenden.²⁵⁾ Überdies sind unsere Analyseprobleme sehr viel heterogener ausgeprägt als etwa zwei Likertskalen-Items, die sich nur in sprachlichen Nuancen unterscheiden. Wegen des breiten Anforderungsspektrums unserer Tests erhöht sich zwar die Validität der Messung, aber man muß mit geringeren Konsistenzkoeffizienten rechnen. Dieser Reliabilitäts-Validitäts-Konflikt ist in der psychometrischen Testtheorie bekannt und von uns zugunsten der Validität entschieden worden, weil man in späteren Phasen der Forschung immer noch die Reliabilität der Messung durch Bildung von Subskalen mit homogenen Falltypen steigern kann. Wir sehen deshalb Werte von Alpha ab 0,60 als für unsere Zwecke zufriedenstellend an.

Um zu überprüfen, welchen Einfluß das in der Schulung angebotene bilanzanalytische Konzept hat, bilden wir Alpha über die experimentellen Fälle 1 bis 4 und über die Fälle 2 bis 4. Außerdem kann überprüft werden, ob die Versuchsbedingung "Klausur" einen Einfluß hat, indem man Alpha über die Fälle 1 bis 4 und den Klausurfall bestimmt. Dies führt zu dem in Tabelle 3.28. dargestellten Befund.

Zwischen den Leistungen eines Analytikers bei verschiedenen Fällen ergeben sich mäßige bis recht starke, signifikant positiv ausgeprägte Korrelationen. Dies weist darauf hin, daß die einzelnen Analysefälle zwar keine inhaltlich gleichen, aber doch formal ähnliche Anforderungen stellen. Betrachtet man alle vier experimentellen Fälle, dann ergeben sich für Cronbachs Alpha Werte, die zwischen 0,50 und 0,77 liegen. Niedrige Werte weisen die Indikatoren Redundanzfreiheit und Qualität der Ursachenanalyse auf. Beim Indikator Redundanzfreiheit ist dies durchaus zu verkraften, denn er stellt eher eine Hilfskennzahl als eine eigenständige Effizienzgröße dar.

Die geringe Konsistenz der Ursachenanalyse kann darauf zurückgeführt werden, daß die sachgerechte Überprüfung fallspezifischer Ursachen offenbar eine stärker fallabhängige Größe ist. Denkbar ist auch, daß die Kreativitätsleistungen, die diesem Indikator zugrundeliegen, nicht beliebig reproduziert werden können. Man kann dies auch an der recht geringen Zahl von sachgerechten "besonderen Verknüpfungen" ablesen, mit denen diese Effi-

Tabelle 3.28.: Konsistenzwerte und Korrelationen der Effizienzindikatoren bei verschiedenen analysierten Fällen

Indikatoren	Korrelationen von/nach			Alpha	Alpha	Alpha
	1-2	2-3	3-4	1-4	2-4	1-K ^{a)}
Treffsicherheit Gesamtlage	0,53	0,46	0,44	0,71	0,69	0,75
Treffsicherheit Erfolgslage	0,38	0,38	0,49	0,68	0,67	0,70
Treffsicherheit Vermögensl.	0,25	0,19	0,39	0,60	0,56	0,51
Treffsicherheit Finanzlage	0,34	0,37	0,23	0,50	0,46	0,52
Urteilsmenge	0,46	0,52	0,65	0,77	0,80	0,76
Qualitatives Niveau	0,21	0,37	0,29	0,62	0,61	0,76
Redundanzfreiheit	0,00	0,27	0,34	0,34	0,49	0,23
Qualität Ursachenanalyse	0,19	0,18	0,22	0,41	0,43	0,47
Allgemeine Frageleistung	0,21	0,41	0,46	0,67	0,68	---b)
Spezifische Frageleistung	0,25	0,32	0,35	0,57	0,52	---b)

a) Die Meßwerte dieser Spalte beruhen auf 64 Analysen pro Fall, alle anderen Meßwerte beruhen auf 103 Analysen pro Fall.

b) Die Frageleistung wurde beim Klausurfall nicht erhoben.

zizienzdimension operationalisiert wird. Die ansteigenden Korrelationskoeffizienten beim Übergang von einem Analysefall zum nächsten deuten jedoch darauf hin, daß es sich um einen Verhaltenszug handelt, der sich zunehmend stabilisiert.²⁶⁾ Dies ist insofern bemerkenswert, weil sich Experten der Bilanzanalyse dadurch auszeichnen, daß sie nach kontradiktorischen Sachverhalten suchen und diesen nachgehen.²⁷⁾

Fazit: Die Konsistenz der interindividuellen Leistungsunterschiede ist recht hoch. Bei dieser Einschätzung muß man berücksichtigen, daß unsere "Items" keine einfachen Ratingfragen darstellen, die nur in Nuancen voneinander abweichen, sondern komplexe Beurteilungsprobleme, die sich inhaltlich recht stark unterscheiden. Wegen ihrer formalen Vergleichbarkeit gibt es aber doch konsistente interindividuelle Leistungsunterschiede. Daher ist es zweckmäßig, einen Durchschnittswert über mehrere Fälle zu bilden, um die Leistung eines Analytikers zuverlässiger zu messen.

3.4.5. Pfadanalyse zur Konstruktvalidierung des Effizienzkonzeptes

Unserer pfadanalytischen Konstruktvalidierung liegen folgende Hypothesen zugrunde:

- (1) Zwischen den Effizienzdimensionen der Analysequalität (Treffsicherheit der Lagebeurteilung, Ursachenanalyse, Frageleistung) besteht ein positiver Zusammenhang.
- (2) Die Effizienzdimensionen der Analysequalität stehen in einem positiven Zusammenhang mit der Humaneffizienz (Zufriedenheit mit Verlauf und Ergebnis der Analyse).
- (3) Die Effizienzdimensionen der Analysequalität stehen in einem positiven Zusammenhang mit der Beurteilungstendenz (Ausprägung der Teilurteile und des Gesamturteils).

Zu (1): Zwischen der Treffsicherheit und der Ursachenanalyse wird eine positive Beziehung vermutet, weil das sachgerechte Erkennen bedeutsamer Stärken und Schwächen eine Voraussetzung für eine vertiefende Ursachenanalyse darstellt. Je intensiver man nach den Ursachen forscht, desto größer ist die Wahrscheinlichkeit, daß man an die Grenzen des Bilanzmaterials stößt und "offene Fragen" formuliert, um die Ursachenforschung mit Hilfe anschließender Informationsaktivitäten fortsetzen zu können.

Zu (2): Den Versuchspersonen wird expressis verbis die Aufgabe vorgegeben, die Lage der Unternehmung zu beurteilen und Fragen für die anschließende Kreditverhandlung zu formulieren. Da sie diese Aufgaben zunehmend besser erfüllten, vermuten wir, daß sie diese Zielsetzungen auch der Beurteilung ihrer eigenen Leistung zugrundelegten. Wir erwarten daher einen positiven Zusammenhang zwischen den objektiv bewerteten Dimensionen der Analysequalität und der subjektiven Zufriedenheit mit der Leistung. Allerdings rechnen wir nur mit einem mäßigen Zusammenhang, weil die erlebten Leistungsverbesserungen auch eine Erhöhung des Anspruchsniveaus bewirken.²⁸⁾

Zu (3): Wer treffsicherer analysiert, mehr Ursachen überprüft und mehr sachgerechte Fragen für die anschließende Kreditverhandlung entwickelt, der sollte auch eine stärker ausgeprägte Beurteilungstendenz besitzen und eher in der Lage sein, ein eindeutiges Urteil zu fällen – zumindest bei den hier vorgegebenen, unterschiedlich gravierend ausgeprägten Fällen. Allerdings wird die Beurteilungstendenz auch durch persönliche Faktoren beeinflusst. Wir gehen daher von einem mäßigen Zusammenhang aus.

Tabelle 3.29. zeigt, daß die Effizienzmeßwerte zur Analysequalität, Beurteilungstendenz und Humaneffizienz untereinander signifikant positiv und hinreichend hoch korrelieren. Alle Korrelationen mit Indikatoren unterschiedlicher Effizienzkonstrukte sind niedriger ausgeprägt. **Die Anforderungen zur Konvergenz- und Diskriminanzvalidität des Systems der Effizienzmeßwerte sind somit erfüllt.**

Zur Überprüfung der nomologischen Validität wird eine Pfadanalyse durchgeführt. Dabei werden die hoch korrelierenden Meßwerte der Humaneffizienz zu einem Faktor zusammengefaßt, weil beide offenkundig dasselbe Konstrukt mes-

Tabelle 3.29.: Korrelationen der Effizienzmeßwerte (Basis: 412 Analysen)

Indikatoren	Analysequalität			Beurteilungstendenz		Humaneffizienz
	Treffsicherheit	Ursachenanalyse	Frageleistung	Tendenz Teilurteile	Tendenz Gesamturteil	Zufriedenheit Verlauf
Ursachenanalyse	0,410					
Frageleistung	0,333	0,361				
Tend. Teilurteile	0,205	0,205	0,135			
Tend. Gesamturteil	0,196	0,165	0,168	0,416		
Zufriedh. Verlauf	0,257	0,065	0,237	0,038	0,083	
Zufriedh. Ergebnis	0,241	0,076	0,196	0,048	0,109	0,810

Abbildung 3.17.: Pfadanalytische Validierung des Effizienzkonzeptes

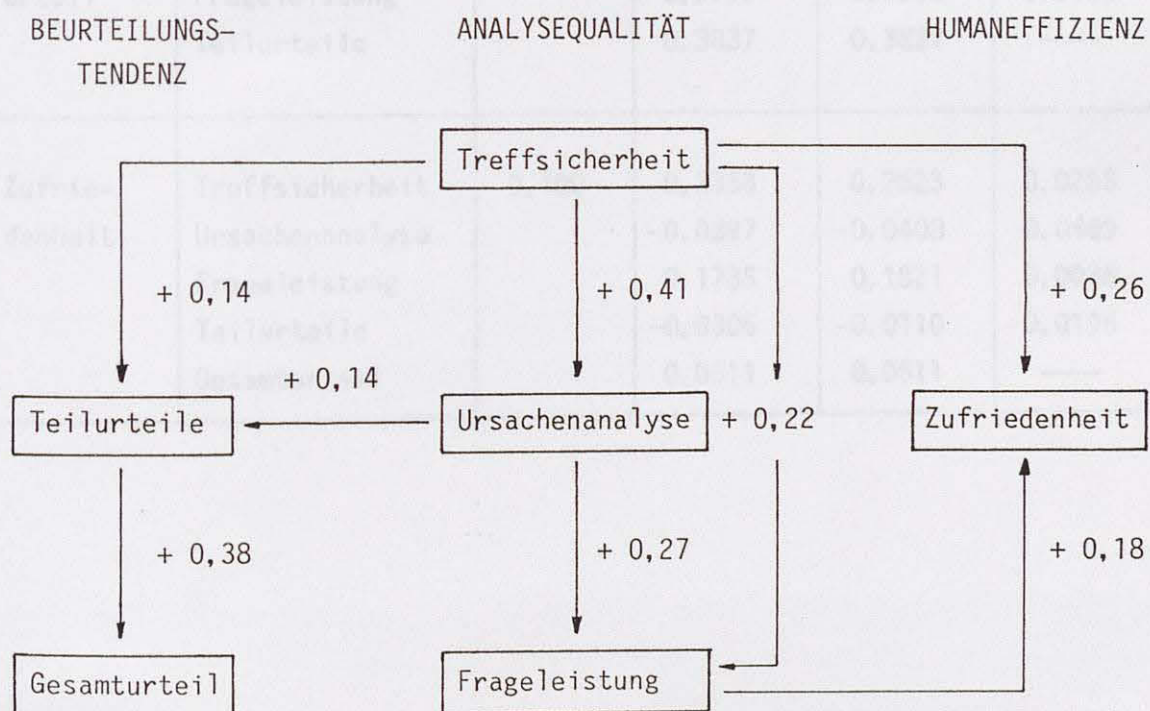


Tabelle 3.30.: Pfadanalyse der Effizienzmeßwerte (Zerlegung der Effekte)

Abhängige Variable	Unabhängige Variablen	Erklärte Varianz	Direkter Effekt	Totaler Effekt	Indirekter Effekt
Ursachenanalyse	Treffsicherheit	0,168	0,4101	0,4101	----
Frageleistung	Treffsicherheit Ursachenanalyse	0,172	0,2230 0,2694	0,3335 0,2694	0,1105 ----
Tendenz Teilurteile	Treffsicherheit Ursachenanalyse Frageleistung	0,061	0,1364 0,1337 0,0417	0,2052 0,1449 0,0417	0,0687 0,0000 ----
Tendenz Gesamturteil	Treffsicherheit Ursachenanalyse Frageleistung Teilurteile	0,193	0,0807 0,0249 0,0796 0,3837	0,1962 0,1020 0,0956 0,3837	0,1155 0,0771 0,0160 ----
Zufriedenheit	Treffsicherheit Ursachenanalyse Frageleistung Teilurteile Gesamturteil	0,100	0,2358 -0,0897 0,1785 -0,0306 0,0511	0,2623 -0,0408 0,1821 -0,0110 0,0511	0,0265 0,0489 0,0036 0,0196 ----

sen und daher eine Faktorenbildung die Zuverlässigkeit der Messung erhöht. Zwischen den beiden Meßwerten zur Frageleistung muß eine Auswahl getroffen werden, weil beide in einer definitorischen Teil-Ganzes-Beziehung stehen. Wir entscheiden uns für die umfassendere Allgemeine Frageleistung, da diese über verschiedene Fälle hinweg eine höhere Konsistenz zeigt. Da die beiden Meßwerte zur Frageleistung hoch miteinander korrelieren, wird das Ergebnis von dieser Auswahl nur unwesentlich beeinflusst.

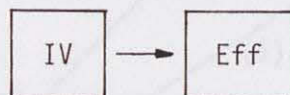
Tabelle 3.30. und Abbildung 3.17. zeigen das Ergebnis unserer Pfadanalyse. (Nur Pfade mit einem Absolutbetrag von 0,10 und mehr werden berücksichtigt).²⁹⁾

Fazit: Der Befund bestätigt alle drei Hypothesen und damit die **nomologische Validität** des Effizienzkonzeptes. Die Aufstellung der direkten, indirekten und totalen pfadanalytischen Effekte zeigt, daß unsere Basisdimension Treffsicherheit der Lagebeurteilung auf alle Effizienzdimensionen einen beachtlichen und in der Richtung hypothesenkonformen Einfluß ausübt. In Verbindung mit den vorangegangenen Befunden können wir daher die **Konstruktvalidität** unseres Meßkonzeptes als bestätigt ansehen.

4.1. THEORETISCHER BEZUGSRAHMEN DER EMPIRISCHEN UNTERSUCHUNG

4.1.1. Das Grundmodell

Gegenstand unserer empirischen Untersuchung ist der Zusammenhang von Informationsverhalten und Effizienz. Unsere Ausgangshypothese lautet: Das Informationsverhalten beeinflusst die Effizienz komplexer betrieblicher Informationsprozesse:



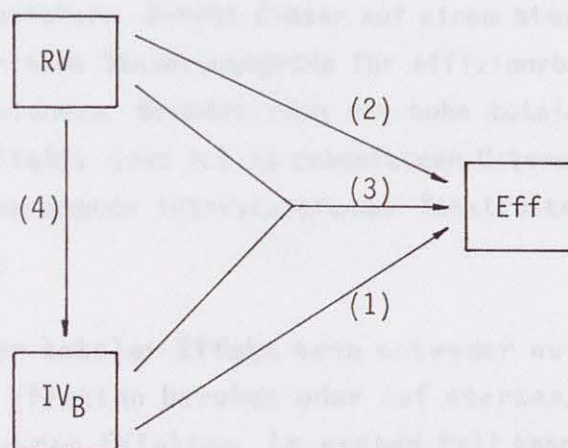
Informationsverhalten ist nicht völlig frei gestaltbar: Seine Ausprägungen und Effizienzwirkungen hängen von den Merkmalen der Aufgabe, des Entscheidungsträgers, des Informationssystems und der Umweltsituation ab. Die meisten experimentellen Effizienzprüfungen gehen von folgender Kausalkette aus: Rahmenvariablen beeinflussen das Informationsverhalten und dieses beeinflusst die Effizienz:



Wie unsere kritische Bestandsaufnahme gezeigt hat, ist ein solches Modell zu einfach. Es unterstellt, daß der Transmissionsriemen "Informationsverhalten" vollständig von den Rahmenvariablen determiniert wird. Tatsächlich gibt es aber innerhalb der gesetzten Rahmenbedingungen noch beachtliche Freiheitsgrade. Ferner dürfte klar sein, daß man die mental ablaufenden Gedächtnisprozesse nur zum Teil aus dem beobachtbaren Informationsverhalten erschließen kann.

Realistischer erscheint daher ein Modell, das sowohl Effizienzeinflüsse der Rahmenvariablen als auch eigenständige Einflüsse des Informationsverhaltens berücksichtigt und zwischen dem beobachteten und dem nicht beobachteten Informationsverhalten unterscheidet (vgl. Abbildung 4.1.).

Abbildung 4.1.: Bezugsrahmen der empirischen Untersuchung



Pfeil (1) beschreibt die direkten Einflüsse des beobachteten Informationsverhaltens auf die Effizienz bei Konstanthaltung der Rahmenvariablen, Pfeil (2) den direkten Effizienzeinfluß der Rahmenvariablen bei Konstanthaltung des beobachteten Informationsverhaltens. Pfeil (3) bezeichnet die Interaktionseffekte von Rahmen- und Verhaltensvariablen. Pfeil (4) drückt die Einflüsse der Rahmenvariablen auf das beobachtete Informationsverhalten aus. Multipliziert man die Effektstärken der Pfeile (1) und (4), erhält man den indirekten Effekt der Rahmenvariablen auf die Effizienz. Direkte und indirekte Effekte addieren sich zum totalen Effekt.

Mit diesem pfadanalytischen Grundmodell¹⁾ lassen sich die Wirkungen des Informationsverhaltens und der Rahmenvariablen differenziert beurteilen.

a) Beurteilung des Informationsverhaltens

Ein **hoher direkter Effekt** wirft die Frage nach den Antrieben, Barrieren und Steuerungsmöglichkeiten des gemessenen Informationsverhaltens auf. Ein **starker Interaktionseffekt** weist auf Kontingenzwirkungen hin, die theoretisch begründet und an neuen Daten überprüft werden sollten. Falls **beide Effekte gering** ausgeprägt sind, wird die Suche nach bisher nicht gemessenen Teilen des Informationsverhaltens vordringlich.

b) Beurteilung der Rahmenvariablen

1. Ein **hoher totaler Effekt** weist darauf hin, daß die Rahmenvariable beachtlich ist und fordert dazu auf, ihren Einfluß in weiteren empirischen Studien abzusichern. Beruht dieser auf einem **starken indirekten Effekt**, dann hat man eine Steuerungsgröße für effizienzbewirkendes Informationsverhalten gefunden. Gründet sich der hohe totale Effekt auf einen **hohen direkten Effekt**, dann ist in zukünftigen Untersuchungen zu klären, welche nicht gemessenen intervenierenden Konstrukte dafür verantwortlich zu machen sind.
2. Ein **geringer totaler Effekt** kann entweder auf geringen direkten und indirekten Effekten beruhen oder auf starken, aber sich gegenseitig kompensierenden Effekten. Im ersten Fall kann die Rahmenvariable vernachlässigt werden, es sei denn, sie weist starke Interaktionseffekte auf. Im zweiten Fall stellt sich die Frage, auf welchen nicht erfaßten Prozessen der direkte Effekt beruht.
3. Ein **starker Interaktionseffekt** weist auf Kontingenzwirkungen hin, die theoretisch begründet und an neuen Daten überprüft werden sollten.
4. Falls **alle drei Effekte gering** sind, kann die Rahmenvariable vernachlässigt werden.

Zum **weiteren Gang der Untersuchung:**

1. Unser besonderes Interesse gilt dem "Pfeil (1)". Es handelt sich dabei nicht um einen einzigen Pfeil, sondern um ein Bündel von Teileinflüssen, die von unterschiedlichen **Dimensionen** des Informationsverhaltens ausgehen. Um diese zu bestimmen, wird zunächst ein Ordnungssystem vorgestellt, das dazu dient, die uns interessierenden **Elemente** des Informationsverhaltens vollständig, überschneidungsfrei, differenziert und reliabel zu messen (Kapitel 4.1.2.). Diese Elemente werden dann zu Dimensionen verdichtet (Kapitel 4.1.3.). In einem weiteren Schritt werden Hypothesen über ihre Zusammenhänge entwickelt (Kapitel 4.1.4.).

2. Aufbauend auf diesen Vorarbeiten werden Hypothesen zur **Effizienzwirkung** der Dimensionen abgeleitet (Kapitel 4.1.5). Sie werden als **direkte** Effekte postuliert, die auf der Prämisse beruhen, daß alle anderen Größen konstant bleiben.
3. Zum Abschluß des Kapitels werden Hypothesen zum Einfluß ausgewählter **Rahmenvariablen** auf das Informationsverhalten und die Effizienz abgeleitet (Kapitel 4.1.6).

4.1.2. Die Elemente des Informationsverhaltens

Unter **Informationsverhalten** verstehen wir mit Witte "... das auf Information gerichtete Tun und Unterlassen von Menschen."²⁾ Es besteht aus verschiedenen Arten von **Informationsaktivitäten**, die sich auf unterschiedliche Gattungen von **Informationsobjekten** beziehen.³⁾ Diese beiden Kategorien sind im folgenden zu differenzieren und auf das untersuchte Problem zu beziehen.

4.1.2.1. Die Informationsaktivitäten

Nach dem sogenannten "Phasen"-Kriterium wird häufig zwischen Aufnahme, Verarbeitung und Abgabe von Informationen unterschieden.⁴⁾ Hinzugefügt wird die auf allen Prozeßstufen auftretende Informationsspeicherung. Außerdem gibt es Meta-Aktivitäten,⁵⁾ die die übrigen Aktivitäten steuern.

Für die vorliegende Effizienzprüfung ist diese Gliederung zu verfeinern und auf **individuelle Analyseprozesse** zu beziehen, die die Aufgabe haben, **Entscheidungen vorzubereiten**. Das Problem sei so gestellt, daß der Aufgabenträger keine Maßnahmen entwickeln, planen, beschließen oder durchsetzen soll. Unser Bezugsrahmen richtet sich ferner auf betriebswirtschaftliche Analyseprobleme, bei denen numerische Informationen zugrundegelegt werden und komplexe Sachverhalte zu beurteilen sind. Die Beurteilung soll sich nicht in einem punktuellen Gesamturteil erschöpfen, sondern auch nach den Ursachen der berichteten Lage forschen, potentielle Risiken und Chancen aufdecken, sowie Fragen aufwerfen, die in weiteren Teilen des (multiperso-

nal) Entscheidungsprozesses beantwortet werden können. Derartige Aufgaben treten nicht nur bei dem hier empirisch untersuchten Problem "Bilanzanalyse in Kreditentscheidungen" auf, sondern auch im Controlling, bei der Frühaufklärung oder bei der Auswertung von empirischen Ergebnissen aus der Marktforschung.

Aus unserer Abgrenzung folgt, daß nur eine Teilmenge des Spektrums an Aktivitäten zu betrachten ist, die bei komplexen, interpersonellen Entscheidungsprozessen auftreten.⁶⁾ Diese Teilmenge ist allerdings differenziert zu beschreiben. Wir unterscheiden folgende **Arten von Informationsaktivitäten**:⁷⁾

- a) Die **Informationsbeschaffung** bezeichnet den Bezug von Informationen von Dritten.⁸⁾
- b) Die **Übertragung und Speicherung** erfaßt das Kopieren von Informationen auf ein externes Speichermedium.
- c) Die **Rechenaktivitäten** messen die numerische Verarbeitung von Informationen. Bei der Bilanzanalyse handelt es sich um die in der Literatur ausführlich erörterte **Bildung von Kennzahlen**.
- d) Die **Beurteilungen** sind wertende Aussagen über betriebswirtschaftliche Sachverhalte. Die Bildung von Urteilen über die wirtschaftliche Lage eines Unternehmens aufgrund von Jahresabschlußinformationen wird im allgemeinen als konstituierendes Element der Bilanzanalyse angesehen.
- e) Unter den **Fragen** sind langschriftlich formulierte Informationsbedürfnisse des Analytikers zu verstehen, die durch die Menge der angebotenen Daten und der aus ihnen vom Analytiker abgeleiteten Informationen nicht befriedigt werden.

Alle aufgeführten Informationsaktivitäten werden in der Literatur zur Bilanzanalyse, auf die sich unser Bezugsrahmen vornehmlich richtet, vorgeschlagen⁹⁾ und mit Effizienzvermutungen belegt. Diese Aktivitäten werden in empirischen Untersuchungen, die das Informationsverhalten bei Kreditent-

4.1. Theoretischer Bezugsrahmen der empirischen Untersuchung Seite - 292 -
scheidungen prozeßbegleitend messen, nachgewiesen.¹⁰⁾ Sie sind in den von uns untersuchten bilanzanalytischen Prozessen großzahlilig aufgetreten¹¹⁾ und konnten mit einer sehr hohen Reliabilität¹²⁾ gemessen werden. Sie sind daher in theoretischer und empirischer Hinsicht als beachtlich anzusehen.

Zur Erklärung der Effizienz werden nur die drei ersten Aktivitätsarten herangezogen, da die Effizienzmessung an den Beurteilungen und Fragen anknüpft. Rein formal zu erklärende, definitorisch-tautologische Zusammenhänge zwischen der Effizienz und den Informationsaktivitäten werden damit ausgeschlossen.

4.1.2.2. Die Informationsobjekte

Die Informationsaktivitäten können sich auf unterschiedliche Gattungen von Informationsobjekten richten. Diese können nach dem Inhalt, dem sachlichen und zeitlichen Bezug sowie dem Aggregationsgrad differenziert werden.

a) Differenzierung der Objekte nach dem Inhalt

In der bilanzanalytischen Literatur werden Aussagen zur Erfolgs-, Vermögens- und Finanzlage unterschieden. Alle drei Bereiche können weiter untergliedert werden.¹³⁾ Unserem theoretischen Modell legen wir nur die primäre Gliederung in Erfolgslage einerseits und Vermögens- und Finanzlage andererseits zugrunde. Es handelt sich hierbei um eine fundamentale Unterscheidung, die sich bereits in den Rechenwerken "Gewinn- und Verlustrechnung" und "Bilanz" des Jahresabschlusses niederschlägt. Sie durchzieht die Kennzahlenvorschläge der bilanzanalytischen Literatur,¹⁴⁾ prägt die Aufbereitungsformulare der Kreditinstitute¹⁵⁾ und beeinflusst das Informationsverhalten der Bilanzleser.¹⁶⁾ Daher wird sie sowohl beim Informationsverhalten als auch bei der Effizienzmessung als Einteilungskriterium verwendet.

b) Differenzierung der Objekte nach dem sachlichen und zeitlichen Bezug

Wer Sachverhalte beurteilen möchte, benötigt dazu Normen und Vergleichsobjekte. Dies gilt nicht nur für pädagogische, juristische und steuerliche Beurteilungsprobleme, sondern auch für Bilanzanalysen. Mit der bilanzanalytischen Literatur unterscheiden wir zwischen Zeit-, Betriebs- und Normvergleichen.¹⁷⁾ Diese Vergleichstechniken werden nicht nur bei der Bilanzanalyse, sondern auch bei anderen betriebswirtschaftlichen Beurteilungsproblemen angewandt.¹⁸⁾

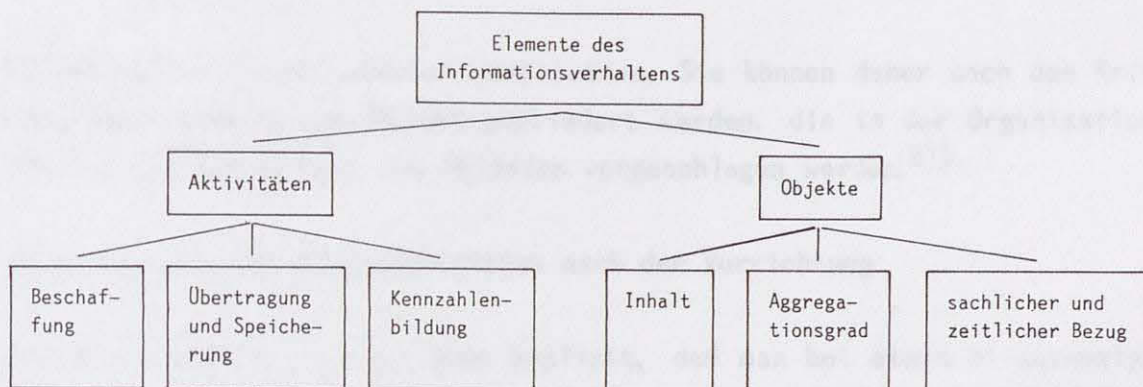
b) Differenzierung der Objekte nach dem Aggregationsgrad

Die Verdichtung von Informationen spielt im Rechnungswesen eine zentrale Rolle. Die bisherige Forschung hat jedoch widersprüchliche Ergebnisse hervorgebracht.¹⁹⁾ Ihr Wert wird stark dadurch beeinträchtigt, daß man nicht gemessen hat, ob und wie die unterschiedlichen Informationsangebote faktisch genutzt werden. Wir wollen daher untersuchen, welche Einflüsse die Beschaffung, Speicherung und Verarbeitung von unverdichteten oder verdichteten Jahresabschlußinformationen auf die Effizienz der Analyse ausübt.

4.1.2.3. Zusammenfassung des Meßkonzeptes

Aus den vorgenommenen Differenzierungen ergeben sich die in Abbildung 4.2. dargestellten Elemente des Informationsverhaltens. Sie können in beliebigen **Kombinationen** auftreten. Bereits bei drei Aktivitätsarten, zwei inhaltlichen Bereichen, zwei Aggregationsgraden und drei Vergleichstechniken würden sich 36 Elementarkombinationen ergeben, zwischen denen 525 paarweise Beziehungen bestünden. Es erscheint jedoch weder theoretisch zweckmäßig noch empirisch realisierbar, auf dieser Datenebene Effizienzhypothesen abzuleiten und zu prüfen. Wir benötigen vielmehr ein theoretisches Konzept, das eine Informationsverdichtung erlaubt.

Abbildung 4.2.: Die Elemente des Informationsverhaltens



4.1.3. Die Dimensionen des Informationsverhaltens

Unser theoretisches Konzept siedelt die Verhaltensdimensionen auf drei verschiedenen Ebenen an:²⁰⁾

- a) Die Verhaltensdimensionen der **strategischen Ebene** beschreiben die Gliederung des Prozesses. Sie geben an, in welcher **Reihenfolge** die einzelnen Teilprobleme abgearbeitet werden.
- b) Die Verhaltensdimensionen der **taktischen Ebene** erfassen die **Struktur** des Prozesses. Sie beziehen sich auf den Anteil der Informationsaktivitäten, der auf bestimmte Teilprobleme entfällt.
- c) Die Verhaltensdimensionen der **operativen Ebene** geben an, mit welchem **Aktivitätsniveau** bestimmte Teilprobleme bearbeitet werden.

Diese rein formale Gliederung nach Reihenfolge, Struktur und Niveau ist inhaltlich zu füllen. Grundsätzlich könnte man zu jedem Problemaspekt Reihenfolge-, Struktur- und Niveauskalen bilden, so daß sich eine Matrix von Verhaltensdimensionen ergibt. Da wir gerade eine Konzentration auf wenige Dimensionen anstreben, wird jeder Problemaspekt nur durch eine Art von Skala erfaßt. Wir nehmen die Zuordnung so vor, daß die grundlegenden Verhaltenselemente durch Reihenfolgeskalen, die sekundären durch Struktur- und die tertiären durch Niveauskalen erfaßt werden.

4.1.3.1. Die strategischen Dimensionen

Bilanzanalysen sind Informationsprozesse. Sie können daher nach den Kriterien **Verrichtung** und **Objekt** gegliedert werden, die in der Organisationstheorie zur Gestaltung von Abläufen vorgeschlagen werden.²¹⁾

a) Gliederung des Analyseprozesses nach der Verrichtung

Bellinger und Sorg postulieren explizit, daß man bei einem bilanzanalytischen Informationsprozeß zuerst Informationen beschaffen, dann bereinigen und aufbereiten, dann zu Kennzahlen verdichten und schließlich beurteilen sollte.²²⁾ Eine derartige "Phasen"-Behauptung wird zu einer empirisch prüf-
baren Existenzhypothese, wenn man sich von dem Gedanken löst, daß eine bestimmte **einzelne** Jahresabschlußinformation immer erst beschafft werden muß, bevor man sie zu Kennzahlen verdichten und beurteilen kann. Wenn man es nämlich mit einer **Vielzahl** von Informationen zu tun hat, kann man bereits bei der zweiten Informationsaktivität entscheiden, ob man die soeben beschaffte Information erst auf ein Formular überträgt oder sie sofort beurteilt, oder ob man als nächstes eine zweite Information beschafft. Ausgehend von dieser Überlegung läßt sich ein Idealtyp der "reinen" oder "strengen" Verrichtungsgliederung konstruieren. Bei der **reinen Verrichtungsgliederung** werden zunächst **alle** Aktivitäten der Art i, dann **alle** Aktivitäten der Art j, dann **alle** Aktivitäten der Art k ausgeführt usw. Aus der Übereinstimmung von empirischer Verrichtungsgliederung und idealer Gliederung kann der Grad der Verrichtungsgliederung abgelesen werden.²³⁾

Wir können hierzu einen von Petersen entwickelten Meßwert übernehmen. Seine Skala zur Verrichtungsgliederung geht von folgender Soll-Reihenfolge der Informationsaktivitäten aus: Beschaffung - Übertragung und Speicherung - Kennzahlenbildung - Beurteilung - Fragen.²⁴⁾ Für alle Paare von artverschiedenen Informationsaktivitäten wird bestimmt, ob ihre zeitliche Aufeinanderfolge der Hypothese entspricht oder nicht. Wenn beispielsweise eine Übertragungs- und Speicherungsaktivität nach einer Beschaffungsaktivität stattfindet, wird eine hypothesenkonforme Proversion gezählt. Bei umgekehrter Reihenfolge wird eine hypothesenkonträre Inversion gezählt.

Als Maßzahl für die "Phasenkonformität der Verrichtungen" wird folgender Index gebildet:²⁵⁾

$$\text{Phasenkonformität der Verrichtungen} = \frac{P - I}{P + I} * 100$$

wobei: P = Anzahl Proversionen, I = Anzahl Inversionen eines Prozesses

Diese Maßzahl kann Werte zwischen - 100 und + 100 annehmen. Ein Wert von + 100 entspricht einer völlig theoremkonformen Sequenz der Aktivitätsarten.

b) Gliederung des Analyseprozesses nach dem Objekt

Für die Objektgliederung bildet Petersen eine analog konstruierte Skala. Bei ihr lautet die Bezugsreihenfolge, daß das Objekt "Erfolgslage" vor dem Objekt "Vermögens- und Finanzlage" bearbeitet wird. Ein Wert von + 100 bedeutet demnach, daß zunächst nur Informationen zur Erfolgslage beschafft und verarbeitet werden. Erst danach richten sich die Informationsaktivitäten auf die Vermögens- und Finanzlage.

4.1.3.2. Die taktischen Dimensionen

Der taktischen Ebene werden zwei Dimensionen zugeordnet: die Komplexität der Vergleichstechnik und der Aggregationsgrad der Informationsaktivitäten.

a) Komplexität der Vergleichstechnik

Bilanzanalytische Urteile können mit Hilfe von Norm-, Zeit- und Betriebsvergleichen abgeleitet werden. Diese Techniken kann man kombinieren, so daß sich eine große Vielfalt möglicher Vergleichstechniken ergibt. Es ist daher nicht einfach, die "Komplexität der Vergleichstechnik" zu messen. In unserem empirischen Feld spitzt sich der Aktionsparameter Vergleichstechnik jedoch auf die Frage zu, ob ein reiner Zeitvergleich oder ein kombinierter

Betriebs- und Zeitvergleich gewählt wird, denn die übrigen Kombinationen treten praktisch nicht auf.²⁶⁾ Zur Operationalisierung der Komplexität der Vergleichstechnik reicht es daher aus, den Anteil der Informationsaktivitäten zu bestimmen, die sich auf das Objekt "Vergleichsunternehmung" beziehen, denn in unserem empirischen Feld gilt die Aussage: Je höher der Anteil an Informationsaktivitäten, die sich auf das Vergleichsunternehmen richten, desto komplexer die Vergleichstechnik.

b) Aggregationsgrad der Informationsaktivitäten

Die Bestimmung des "Aggregationsgrades" von Informationen stellt ein außerordentlich schwieriges, bisher noch ungelöstes Problem dar. Formale Einteilungskriterien helfen nur wenig: Sicherlich ist das "Anlagevermögen" eine stärker aggregierte Größe als das Sachanlagevermögen, weil die zweite Position eine Teilmenge der ersten darstellt. Doch wie soll man beispielsweise die bereits im Rechnungswesen hoch aggregierte Größe "Umsatz" einordnen? Soll sie eine "disaggregierte" Position darstellen, nur weil es keine Summenposition mehr "über" ihr gibt?²⁷⁾

Wir haben uns aus diesen grundsätzlichen Schwierigkeiten für eine pragmatische Lösung entschieden. Der "Aggregationsgrad" drückt aus, welcher Anteil der Informationsaktivitäten auf die zusätzlich angebotenen Summen- und Saldenpositionen entfällt. Diese Meßvorschrift ist nicht nur leicht anwendbar und zuverlässig, sie läßt vor allem klar erkennen, welche Effizienzwirkungen die Nutzung einer derartigen Informationshilfe auslöst.

4.1.3.3. Die operativen Dimensionen

Den operativen Dimensionen werden die Aktivitäten Informationsbeschaffung, Übertragung und Kennzahlenbildung zugeordnet:

- * Die Effizienzwirkungen der **Informationsbeschaffung** werden besonders häufig untersucht. Sie belegen einen positiven Einfluß (vgl. Kapitel 2.4.). Die auch als "aquisition", "Aufnahme", "demand", "Nachfrage", "processing", "request", "retrieval", "R-Nutzung", "Sammlung", "scanning",

"search", "screening", "Suche", "tracking", "usage" bezeichnete Informationsaktivität darf deshalb in unserem Modell nicht fehlen.

- * Ein häufig empfohlener Vorgang bei der Bilanzanalyse stellt das **Übertragen** erhaltener Informationen in sogenannte Bilanzgliederungsformulare dar.²⁸⁾ Wir simulieren diesen Vorgang, indem die experimentellen Daten in Form vieler kleiner Informationskärtchen angeboten werden, die in Taschen großer Informations-Display-Boards stecken, so daß sich eine gewisse Notwendigkeit ergibt, die Daten auf einem externen Speicher in einer übersichtlichen Form anzuordnen (vgl. Kapitel 4.2.).
- * Daß zur Bilanzanalyse **Kennzahlenbildung** gehört, wird von keinem Autor bestritten. Es gibt jedoch durchaus Hinweise auf dysfunktionale Wirkungen einer "zu großen" Menge an Kennzahlen. Um diese Vermutungen empirisch zu prüfen, nehmen wir auch diese Verhaltensdimension in unser Modell auf.

4.1.4. Hypothesen zum Zusammenhang der Dimensionen

4.1.4.1. Abgrenzung

Unsere Analyse der Zusammenhänge zwischen den Dimensionen des Informationsverhaltens beschränkt sich auf ausgewählte Beziehungen. Hierfür sind sowohl inhaltliche als auch methodische Überlegungen maßgeblich:

1. Unser Modell ist so konzipiert, daß die einzelnen Verhaltensdimensionen Aktionsparameter darstellen, die man auf verschiedene Art und Weise kombinieren kann, um eine höhere Effizienz des Informationsprozesses zu erreichen.
2. Die empirische Forschung belegt große Unterschiede im Informationsverhalten. Sie weist darauf hin, daß die Kombinationsmöglichkeiten nicht nur theoretisch vorhanden, sondern auch empirisch genutzt werden. Auf diesem Befund baut die Forschung über Kognitive Stile auf. Sie versucht gerade diese Unterschiede durch Persönlichkeitsmerkmale zu erklären.

3. Die Ebenen unseres Modells erfassen die Reihenfolge, die Struktur und das Niveau des Informationsverhaltens. Zwischen diesen Aspekten ergeben sich bereits dann deutlich unter Eins liegende Korrelationen, wenn man sie auf denselben Inhalt bezieht. So korreliert die Skala zur Reihenfolge der Bearbeitung der Analysefelder "Erfolgslage" und "Vermögens- und Finanzlage" mit dem Anteil Informationsaktivität, der sich auf die Erfolgslage richtet, mit 0,425. Dies entspricht einer Varianzerklärung von 18,1 v. H. Vergleichbare Korrelationskoeffizienten aus der Marketingforschung liegen zwischen 0,20 und 0,60.²⁹⁾ Da sich die Verhaltensdimensionen unseres Modells jedoch nicht auf inhaltlich gleiche, sondern auf unterschiedliche Aspekte beziehen, rechnen wir mit deutlich geringeren Zusammenhängen.

Aufgrund dieser Überlegungen beschränken wir uns auf Zusammenhänge zwischen den Verhaltensdimensionen gleicher Ebenen.

4.1.4.2. Zusammenhänge zwischen den strategischen Dimensionen

Die Zusammenhänge zwischen den strategischen Dimensionen werden in der Arbeit von Petersen untersucht und sind daher nicht Gegenstand unserer Schrift.³⁰⁾ Wir weisen nur darauf hin, daß sich lediglich eine **strenge** Gliederung nach Verrichtungen und eine strenge Gliederung nach Objekten wechselseitig ausschließen. Dazwischen gibt es eine Reihe von Kombinationsmöglichkeiten, die von Petersen systematisch behandelt werden.

4.1.4.3. Zusammenhänge zwischen den taktischen Dimensionen

Zwischen der Komplexität der Vergleichstechnik und dem Streben nach Detailinformation ergibt sich bei gegebenem Zeitbudget ein **Kapazitätskonflikt**: Man kann nicht gleichzeitig in die Tiefe gehen und bei **allen** Informationen einen kombinierten Zeit- und Betriebsvergleich durchführen. Man muß vielmehr abwägen, ob man auf einer stärker aggregierten Ebene verharret und diese Informationen durch komplexere Vergleichstechniken umsichtiger

4.1. Theoretischer Bezugsrahmen der empirischen Untersuchung Seite – 300 –
würdigt, oder ob man eine Kennzahl in Unterkennzahlen und Einzelpositionen zerlegt und diese lediglich im Zeitvergleich beurteilt.

Es ist möglich, daß die Analytiker diesen Kapazitätskonflikt **nicht erkennen** und sowohl zahlreiche detaillierte Informationen als auch zahlreiche Vergleichsinformationen einholen, aber es erscheint wegen der begrenzten Analysezeit unwahrscheinlich, daß sie diese Strategie bei der Übertragung und Kennzahlenbildung fortsetzen können. Da sich unsere Maßzahlen auf alle drei Aktivitätsarten gründen, gehen wir von einer konfliktären Beziehung zwischen beiden Verhaltensdimensionen aus:

Je höher die Komplexität der Vergleichstechnik, desto höher der Aggregationsgrad der Informationsaktivitäten.

4.1.4.4. Zusammenhänge zwischen den operativen Dimensionen

Mit zunehmender Menge an beschafften Informationen ergibt sich für den Analytiker die Notwendigkeit, die Informationen auf einem externen Speichermedium übersichtlich anzuordnen. Je mehr Informationen beschafft und übertragen werden, desto größer ist der Bedarf an Kennzahlen, die diese Informationen verdichten. Auch wenn man die Beeinflussungsrichtung umkehrt, läßt sich ein positiver Zusammenhang postulieren: Weil mehr Kennzahlen gebildet werden sollen, beschafft der Analytiker mehr Daten.³¹⁾ Weil er den Analysebogen als übersichtliches Rechen- und Kontrollblatt benutzt, überträgt er bei einer größeren Zahl von Kennzahlen auch mehr Daten auf den Analysebogen.

Unabhängig davon, welche Beeinflussungsrichtung gilt, lassen sich diese Überlegungen zu folgenden Hypothesen zusammenfassen:³²⁾

Je mehr Informationen beschafft werden, desto mehr Informationen werden auf den Analysebogen übertragen.

Je mehr Informationen beschafft werden, desto mehr Kennzahlen werden gebildet.

Je mehr Informationen übertragen werden, desto mehr Kennzahlen werden gebildet.

Dieses Modell vernachlässigt die **Kapazitätskonkurrenz** zwischen den Informationsaktivitäten wegen der gegebenen Analysezeit: Ein höheres Niveau einer Aktivitätsart verringert das Niveau einer anderen Art. Diese Hypothese wird durch die empirischen Untersuchungen von Schroder, Driver und Streufert gestützt. Sie stellen eine umgekehrt u-förmige Beziehung zwischen unterschiedlich hohen, sequentiell verabreichten Liefermengen und verschiedenen Meßwerten des Niveaus der Informationsintegration fest.³³⁾ Es lassen sich demnach folgende Alternativhypothesen aufstellen:

Mit zunehmender Anzahl beschaffter Informationen nimmt die Anzahl übertragener Informationen zunächst zu und dann ab.

Mit zunehmender Anzahl beschaffter Informationen nimmt die Anzahl gebildeter Kennzahlen zunächst zu und dann ab.

Mit zunehmender Anzahl übertragener Informationen nimmt die Anzahl gebildeter Kennzahlen zunächst zu und dann ab.

4.1.5. Hypothesen zur Effizienzwirkung des Informationsverhaltens

Die folgenden Hypothesen beziehen sich auf die Effizienzdimension "Nutzen der Analyse", die im dritten Teil dieser Arbeit operationalisiert wird.

4.1.5.1. Effizienzwirkungen der Informationsbeschaffung

Zwischen der Menge an beschafften Informationen und der Effizienz der Analyse wird eine **umgekehrt u-förmige Beziehung** vermutet.

Der "aufsteigende Ast" dieser Funktion kann wie folgt begründet werden: Ohne Informationsbeschaffung können keine sachlich zutreffenden Urteile abgeleitet werden. Eine Mindestmenge an Informations-"Bausteinen" ist erforderlich, um Kennzahlen zu bilden und Urteile zu begründen.

Für den "absteigenden Ast" sind folgende Gründe verantwortlich zu machen:

1. Mit zunehmender Anzahl an beschafften Informationen wächst die **Gefahr der Wissensduplizierung**:³⁴⁾ Die zusätzlich beschafften Bausteine sind nur noch bedingt geeignet, neue Schlüsse abzuleiten.³⁵⁾
2. Eine wachsende Menge beschaffter Informationen führt zu einer **Einschränkung der Informationsverarbeitung**. Der negative Effizienzeinfluß äußert sich nicht nur in der bereits behandelten Reduzierung der Kennzahlenberechnung. Auch die (sachgerechte) **Interpretation** der Kennzahlen wird behindert.
3. In der Fülle der beschafften Daten werden die wesentlichen Informationen übersehen. Es ergibt sich ein dysfunktionaler **Ablenkungseffekt**.³⁶⁾
4. Wenn zu viele Daten beschafft und mental aufgenommen werden, treten **negative Interferenzeffekte** auf: Bereits gebildete, aber noch nicht verfestigte Lösungshypothesen werden durch nachdrängende Reize überschneidet und zerfallen.³⁷⁾

Wir formulieren daher folgende Hypothese:

Mit zunehmender Anzahl beschaffter Informationen nimmt die Effizienz der Analyse zunächst zu und dann ab.

Diese Hypothese beschreibt keine naturgesetzliche Zwangsläufigkeit, sondern mögliche dysfunktionale Effekte. Wir gehen davon aus, daß diese mit zunehmender Erfahrung seltener auftreten: Erfahrene Bilanzanalytiker können wegen ihres größeren Vorwissens Informationen selektiver nachfragen. Sie besitzen eine größere "Chunking"-Fähigkeit, bestimmte Muster zu erkennen und auf Schlüsselinformationen zu reduzieren.³⁸⁾ Schließlich haben sie gelernt,

4.1. Theoretischer Bezugsrahmen der empirischen Untersuchung Seite - 303 -
mit ihren persönlichen Informationsverarbeitungsfähigkeiten umzugehen und stimmen ihr Verhalten darauf ab. Deshalb postulieren wir folgenden Interaktionseffekt:

Bei ungeübten Analytikern treten eher dysfunktionale Effekte einer zu hohen Beschaffungsmenge auf als bei geübten Analytikern.

4.1.5.2. Effizienzwirkungen der Übertragung und Speicherung

Eine Übertragung der Ausgangsdaten auf ein Formular gehört zu den Routineaktivitäten der Bilanzanalyse in vielen Kreditinstituten. Durch diesen Vorgang werden die heterogen gegliederten Bilanzen der Kunden aus unterschiedlichen Branchen und Rechtsformen homogenisiert und auf ein vergleichbares Standardschema verdichtet.³⁹⁾

In unseren Experimenten greift diese Effizienzerklärung nicht, da es sich bereits um einheitlich gegliederte Daten handelt. Trotzdem gibt es auch hier gewisse Rationalisierungs- und Kreativitätseffekte einer Übertragung der Daten von den angebotenen Informationskärtchen auf die Analysebögen in Form von Tabellen, Staffeln, Bäumen oder Konten:

- * Eine übersichtliche, sachlich zweckmäßige Anordnung der Ausgangsdaten und Kennzahlen erleichtert die systematische Zerlegung von Kennzahlen und die Durchführung von Vergleichen.
- * Ein zentrales Speichern häufig benötigter Daten wie z. B. Bilanzsumme, Umsatz oder Gesamtleistung erleichtert das Wiederauffinden und beschleunigt die Kennzahlenbildung.
- * Rechenfehler können leichter festgestellt und korrigiert werden, wenn man die zugehörigen Daten und Zwischenergebnisse auf einem Blatt abspeichert.

- * Formulare verhindern, daß man vergißt, wichtige Informationen zu prüfen.
- * Schließlich wird vielfach vermutet, daß man sich beim Übertragen der Daten diese auch mental besser einprägt.

Auf der anderen Seite ist das Übertragen unbearbeiteter Daten ein zeitaufwendiger Vorgang, der nur einen geringen Erkenntnisfortschritt bewirkt. Die verbrauchte Kapazität fehlt bei der Kennzahlenbildung und Beurteilung.

Wir vermuten, daß es eine optimale Übertragungsmenge gibt, die sich darin äußert, daß man sich auf wichtige Positionen beschränkt und diese übersichtlich anordnet. Wir postulieren daher:

Mit zunehmender Anzahl übertragener Informationen nimmt die Effizienz der Analyse zunächst zu und dann ab.

4.1.5.3. Effizienzwirkungen der Kennzahlenbildung

"Kennzahlen ... gehören bereits seit den Anfängen des Einsatzes wissenschaftlicher Methoden der Betriebswirtschaftslehre in der Praxis zu den wichtigsten Instrumenten einer rationalen Entscheidungsfindung. Ohne Übertreibung läßt sich deshalb feststellen, daß in der einem ständigen Wandel unterworfenen Unternehmens- und Umweltsituation die von einem Entscheidungsträger laufend zu erstellenden Analysen und zu treffenden Entscheidungen ohne den Einsatz von Kennzahlen kaum zu bewältigen sind."⁴⁰⁾

"Die Auswahl, die Errechnung und der Vergleich von Kennzahlen stellt den **Kern der Bilanzanalyse** dar. Es fällt nicht schwer, **Bilanzanalyse überhaupt als Kennzahlenberechnung zu bezeichnen**, wenn auch die mit Hilfe von Kennzahlen gewonnenen Erkenntnisse durch andere zusätzliche Einsichten und Informationen ergänzt und abgerundet werden."⁴¹⁾

Worauf stützt sich die hinter solchen Aussagen steckende Effizienzvermutung? Unseres Erachtens erwartet man von Kennzahlen, insbesondere in sinnvoll geordneten Systemen, folgende Leistungen:⁴²⁾

1. **Verdichtungsfunktion:** Kennzahlen reduzieren ein komplexes Zahlenwerk auf wenige, überschaubare Aggregatgrößen. Sie lenken die Aufmerksamkeit auf "das Wesentliche" und kommen der beschränkten Verarbeitungskapazität des Menschen entgegen.
2. **Ordnungsfunktion:** Kennzahlen gliedern Gesamtmassen in Teilmassen, bündeln homogene Gruppen, sondern heterogene aus, heben "harte" von "weichen" Daten ab. Sie fördern damit eine Problemlöseaktivität, die von dem Psychologen Dörner "Komponentenanalyse" genannt wird und darin besteht, daß man komplexe Sachverhalte erschließt, indem man Teil-Ganzes-Relationen herstellt.⁴³⁾
3. **Relativierungsfunktion:** Kennzahlen machen aus schlecht vergleichbaren Rohdaten besser vergleichbare "Verhältnis"-Zahlen. So kann man beispielsweise den unvergleichbaren Gewinn von Betrieben unterschiedlicher Größe in eine vergleichbare Renditegröße umwandeln, indem man ihn an einem geeigneten Größenmaß relativiert.
4. **Erkenntnisfunktion:** Kennzahlen geben als Indexzahlen Aufschlüsse über zeitliche Entwicklungen.⁴⁴⁾ Als Beziehungszahlen verknüpfen sie (vermutete) Ursachen und Wirkungen und lassen Ausmaß und Veränderung wichtiger Zusammenhänge erkennen. Sie unterstützen damit die erkenntnistiftenden Problemlösungsaktivitäten Trend- und Dependenzanalyse.⁴⁵⁾

Es verwundert daher nicht, daß bei der Bilanzanalyse eine Kennzahlenbildung empfohlen wird. Betrachtet man die Fülle der vorgeschlagenen Kennzahlen, dann kann man den Eindruck gewinnen, daß bereits zwischen der **Kennzahlenberechnung** und der Analysequalität eine positive Beziehung vermutet wird.

Doch vor einer mechanistischen Interpretation des behaupteten positiven Zusammenhangs zwischen Kennzahlenberechnung und Effizienz muß gewarnt werden: Mangelnde Sachgerechtigkeit, geringe, selten überprüfte empirische Relevanz und fehlende Überschneidungsfreiheit der häufig unverbindlich aneinandergereihten "Kenn"-Zahlen beschwören die Gefahr, daß Datentabellen lediglich durch Kennzahlenansammlungen ersetzt werden.

Auf diese Gefahren wird in der einschlägigen Literatur durchaus hingewiesen: Merkle warnt vor einer "Kennzahleninflation",⁴⁶⁾ Küting fordert ein "kompaktes System",⁴⁷⁾ Reichmann und Lachnit eine Beschränkung auf "das Wesentliche",⁴⁸⁾ Hauschildt und Staudt et al. sehen den Analytiker gar in einem "Kennzahlendschungel".⁴⁹⁾ Die Literatur sieht eine Lösung dieses Problems vor allem in konsistenten, aufeinander abgestimmten Kennzahlensystemen, die je nach individuellem Bedarf feiner oder gröber gestaltet werden können.⁵⁰⁾

In unseren Experimenten wird den Analytikern ein empirisch validiertes System angeboten, das dem "Du-Pont-Control-System"⁵¹⁾ nachempfunden ist und sich auf eine vergleichsweise geringe Zahl⁵²⁾ von Kennzahlen beschränkt.

Wir könnten daher von einer positiven Beziehung zwischen dem Aktivitätsniveau der Kennzahlenberechnung und der Effizienz ausgehen. Auf der anderen Seite kann nicht ausgeschlossen werden, daß ungeübte Analytiker schlechten Beispielen der bilanzanalytischen Literatur folgen und mehr Kennzahlen als erforderlich bilden. Aus Unsicherheit und Abneigung gegenüber einem klaren Urteil werden Kennzahlen berechnet, **anstatt** Sachverhalte zu beurteilen. Wir stellen daher folgende Effizienzhypothese auf:

Mit zunehmendem Aktivitätsniveau der Kennzahlenbildung nimmt die Effizienz der Analyse zunächst zu und dann ab.

4.1.5.4. Effizienzwirkungen des Aggregationsgrades

Bezüglich der Nutzung der zusätzlich angebotenen Kennzahlen können wir idealtypisch drei Vorgehensweisen unterscheiden:⁵³⁾

1. **Reiner "Bottom-Up-Approach":** Der Analytiker verzichtet völlig auf die zusätzlich angebotenen verdichteten Informationen. Er bildet stattdessen eigene Summen- und Saldenpositionen und geht dabei von "unten nach oben vor".

2. Reiner "Top-Down-Approach": Der Analytiker orientiert sich nur an den zusätzlich angebotenen Summen- und Saldenpositionen. Er geht dabei von "oben nach unten vor".

3. Gemischte Strategie: Der Analytiker geht sowohl von "unten nach oben" als auch von "oben nach unten" vor. Er wechselt überdies zwischen den Rechenwerken. Damit kann er die volle Breite und Tiefe des Informationsangebots ausschöpfen. Wir vermuten jedoch, daß er dies nicht tut, sondern mit stärker aggregierten Summen- und Saldenpositionen beginnt und diese je nach Ausprägung des konkreten Falls weiter zerlegt und mit Informationen aus anderen Rechenwerken verknüpft.

Wir vermuten, daß die beiden reinen Typen zu einer geringen Effizienz der Analyse führen:

- * Wer völlig auf die Nutzung sachgerecht aufbereiteter Informationen verzichtet, muß alle aggregierten Informationen selbst erzeugen. Dieser Vorgang ist fehleranfällig. Die verbrauchte Zeit fehlt bei der Beurteilung der gebildeten Kennzahlen.
- * Wer nur aufbereitete Daten beurteilt, läuft Gefahr, kompensierende Saldierungen von Teileinflüssen zu übersehen. Er erhält nur wenige Hinweise auf die Ursachen der beobachteten Entwicklung oder versäumt es, seine diesbezüglichen Hypothesen zu prüfen.

Eine integrative Verknüpfung von aufbereiteten und primären Daten vermeidet diese Irrtümer. Die Ursachenanalyse sollte aber nicht nur in den vorgezeichneten Bahnen des Kennzahlenbaums verlaufen, denn mit der reinen " Lokalisation der Veränderungen werden ... die letzten Ursachen nicht erkennbar."⁵⁴⁾ Es kommt vielmehr darauf an, sachlogisch begründeten Beziehungen zu anderen Ästen des Baumes nachzuspüren oder diesen ganz zu verlassen. Im Zweifel wird nur der Analytiker mit einem mittleren Anteil aufbereiteter Kennzahlen diese Fähigkeit besitzen. Wir postulieren daher:

Mit zunehmendem Aggregationsgrad der Informationsaktivitäten nimmt die Effizienz der Analyse zunächst zu und dann ab.

4.1.5.5. Effizienzwirkungen der Vergleichstechnik

Auch bezüglich der Komplexität der Vergleichstechnik postulieren wir eine umgekehrt u-förmige Beziehung:

- * Analytiker, die nur Zeitvergleiche anstellen, laufen Gefahr, die beobachteten Veränderungen unangemessen stark zu bewerten. Praktiker der Unternehmensanalyse heben hervor, daß man nicht auf einen Betriebsvergleich verzichten solle.⁵⁵⁾ Unser empirischer Test weist nach, daß der Betriebsvergleich bei vielen Urteilsbereichen Problem- und Vergleichsfälle besser trennt als der Zeitvergleich (vgl. Kapitel 3.2.3.2.).
- * Auf der anderen Seite muß vor einer schematischen Anwendung des Betriebsvergleichs gewarnt werden: So weist bereits Schnettler auf vielfältige "Störfaktoren" hin.⁵⁶⁾ Auch wenn man Hauschildt zustimmen muß, daß man nicht trotz, sondern gerade wegen dieser "Störfaktoren" angeregt wird, Fragen zu stellen, ob denn tatsächlich bestimmte Faktoren Unterschiede in der wirtschaftlichen Leistung hervorgerufen haben,⁵⁷⁾ kann dies nicht heißen, daß solche Fragen zu jeder einzelnen Position aufgeworfen werden.

Wir formulieren daher:

Mit zunehmender Komplexität der Vergleichstechnik nimmt die Effizienz der Analyse zunächst zu und dann ab.

4.1.5.6. Effizienzwirkungen der Abfolge der Verrichtungen

1. Eine **hohe Verrichtungsgliederung** birgt die Gefahr, daß man viele Informationen beschafft und aufbereitet, die sich beim konkreten Fall als irrelevant herausstellen. Die verbrauchte Kapazität fehlt bei der Beurteilung, Ursachenforschung und Fragenentwicklung.

2. Bei einer **niedrigen Verrichtungsgliederung** werden ganz kleine Informationspäckchen beschafft, übertragen und sofort zu Kennzahlen und Urteilen verdichtet. Es besteht die Gefahr, daß man die einzelnen Aspekte isoliert betrachtet, sich stark von "Auffälligkeiten" des Einzelfalls leiten läßt und die integrierende Gesamtschau für das Ganze verliert. Wir vermuten, daß eine solche Vorgehensweise vor allem von **ungeübten Analytikern** gewählt wird, die noch nicht über "Mustererkennungs"-Routinen verfügen.

Nach Simon unterscheiden sich Laien von Experten durch die Komplexität der in ihrem Langzeitgedächtnis gespeicherten Informationen zu einer bestimmten "Domäne".⁵⁸⁾ So besteht die Stellung eines Schachspiels für einen Laien aus 25 Einzelfiguren, von denen er sich aber nur ca. 5 bis 9 in seinem Kurzzeitgedächtnis merken kann.⁵⁹⁾ Für den Großmeister reduziert sie sich auf ein Muster von ca. 7 **Gruppen** mit je drei bis fünf Figuren, die er ohne Mühe in seinem Kurzzeitgedächtnis verfügbar halten kann. Diese "Chunking"-Fähigkeit zeichnet nicht nur geübte Schachspieler, sondern auch Experten auf anderen "semantisch reichen Domänen" aus, z. B. Ingenieure, Ärzte, Juristen, Wirtschaftsprüfer und Manager. Simons Thesen werden gestützt durch neuere Ergebnisse der psychologischen Problemlöseforschung.⁶⁰⁾ Bouwman belegt solche Unterschiede für die Bilanzanalyse.⁶¹⁾

3. Wir vermuten, daß ein **mittlerer Grad der Verrichtungsgliederung** einer modularen, relevanzgesteuerten Bearbeitung von Teilproblemen entspricht. Bei der Bilanzanalyse entspricht dies einer Zerlegung in bestimmte Felder, die er nacheinander und in **Abhängigkeit** von den fallspezifischen Beobachtungen überprüft. Er geht nicht nach einem starren Plan vor, der in einem Formular festgeschrieben ist. Wir wollen eine solche Vorgehensweise an einem Beispiel veranschaulichen:

Der Analytiker zerlegt das Gesamtproblem "Bilanzanalyse" in die Teilprobleme "Analyse der Erfolgslage" und "Analyse der Vermögens- und Finanzlage". Innerhalb der "Erfolgslage" bildet er die Analysemodule "Gesamterfolg", "Quellen des Gesamterfolgs", "Quellen der Teilerfolge". Er beginnt mit dem Gesamterfolg und vergleicht dessen Höhe und Entwicklung mit der seiner Quellen. Erst dann **entscheidet er**, und zwar in Abhängigkeit von diesem Vergleich, in welcher Reihenfolge und Intensität **deren**

Quellen untersucht werden sollen. Er wendet sich erst dann der Beschaffung, Aufbereitung und Verdichtung der Informationen der **fallspezifisch** relevanten Quellen zu. Dies bedeutet, daß er die irrelevant erscheinenden Module vernachlässigen und die als wichtig erkannten vertiefen kann.

Wir vermuten, daß diese Verhaltensweise eine besonders wirksame Koordination der Informationsaktivitäten ermöglicht. Die Gefahr einer "unfertigen Analyse", die lediglich aufbereitete Kennzahlen, aber noch keine Urteile produziert, wird vermieden. Wir stellen daher folgende Hypothese auf:

Mit zunehmendem Grad der Verrichtungsgliederung der bilanzanalytischen Aktivitäten nimmt die Effizienz der Analyse zunächst zu und dann ab.

4.1.5.7. Effizienzwirkungen der Abfolge der Objekte

Wie liest man eine Bilanz? In welcher **Reihenfolge** soll man die Analysefelder der Erfolgs-, Vermögens- und Finanzlage bearbeiten, um eine möglichst hohe Effizienz der Analyse zu erreichen?

Die bilanzanalytischen Lehrbücher⁶²⁾ geben selten eine Antwort auf diese Frage: Nur Hauschildt und Gräfer äußern sich explizit. Beide plädieren dafür, die Erfolgslage **vor** der Vermögens- und Finanzlage zu analysieren.⁶³⁾

Wir halten die Tatsache, daß man auf dieses wichtige Organisationsproblem in Lehrbüchern zur Bilanzanalyse, die ja vor allem **ungeübten** Analytikern **praktische** Hinweise zur Durchführung von Bilanzanalysen geben wollen, für ein **gravierendes Defizit**: Die menschliche Analysekapazität ist ein knappes und teures Gut. Es ist daher ein beachtliches ökonomisches Problem, welche Teile eines Jahresabschlusses zuerst analysiert werden. Man sollte erwarten, daß die wichtigeren Aspekte zuerst geprüft werden. Man sollte anerkennen, daß die Analyse nachfolgender Bereiche von dem Urteil abhängt, das man sich bei den zuvor bearbeiteten Bereichen gebildet hat. So wird man bei der wichtigen Frage der Bewertung von Aktiva und Passiva bei einer guten Erfolgslage eine ganz andere bilanzpolitische Hypothese vertreten als bei einer schlechten Erfolgslage.

Denkbar ist jedoch, daß die bilanzanalytische Literatur **implizit** eine bestimmte Vorgehensweise empfiehlt und dabei eine so **einheitliche Meinung** vertritt, daß das Reihenfolgeproblem als "gelöst" anzusehen ist.

Um diese Meinung zu ergründen, gehen wir von einem ungeübten Analytiker aus, der vor dem Problem steht, seine Analyse nach bestimmten Inhalten zu gliedern. Er wendet sich an die bilanzanalytische Literatur und gliedert seine Analyse nach der Reihenfolge, in der man ihm Kennzahlen zur Erfolgs-, Vermögens- und Finanzlage präsentiert. Unsere Ausgangsfrage lautet demnach: **Welcher Prozeßverlauf ergibt sich, wenn man die Analysefelder in der Reihenfolge bearbeitet, die in bilanzanalytischen Monographien angewandt wird?**

Um diese Frage zu beantworten, wird eine Prozeßanalyse von 24 bilanzanalytischen Monographien vorgenommen, die systematische Anleitungen geben, welche Kennzahlen man bilden sollte. Es handelt sich um die gleichen Monographien und Kennzahlen, die wir bereits bei der Bestimmung des "literarischen Relevanzmaßes" herangezogen haben. Diese Werke werden nach folgenden Kriterien quantitativ ausgewertet:

- a) **Menge:** Anzahl der gebildeten Kennzahlen;
- b) **Inhalt:** Anteil der Kennzahlen zur Erfolgslage, in v. H. aller gebildeten Kennzahlen;
- c) **Reihenfolge:**
 - 1. Anzahl der Kennzahlen zur Erfolgslage innerhalb der ersten 10 Kennzahlenvorschläge,
 - 2. Wechsel zwischen den Analysefeldern, beim Übergang von einer Kennzahl zur nächsten (Transitionsanalyse),
 - 3. Indexwert der Objektgliederung in Richtung "Erfolgslage" vor "Vermögens- und Finanzlage".

Die Transitionsanalyse betrachtet die Übergänge von einer Kennzahl zur unmittelbar nachfolgenden.⁶⁴⁾ Es bezeichne "E" eine Kennzahl zur Erfolgslage und "VF" eine Kennzahl zur Vermögens- und Finanzlage. Man kann vier

mögliche Transitionen unterscheiden:

1. "E-E",
2. "E-VF",
3. "VF-E" und
4. "VF-VF".

Uns interessieren die **Wechsel** zwischen den Urteilsbereichen. Wir bestimmen deshalb den Anteil der Folgen "E-VF" und "VF-E" an der Summe aller Transitionen erster Ordnung.

Der Index zur Objektgliederung stellt ein feineres Maß dar. Er berücksichtigt die Beziehungen zu **allen** Nachfolgern, nicht nur zur unmittelbar nachfolgenden Kennzahl. Er ist definiert als:

$$\text{Index der Objektgliederung "Erfolgslage" vor "Vermögens- und Finanzlage"} = \frac{\text{E vor VF} - \text{VF vor E}}{\text{E vor VF} + \text{VF vor E}}$$

dabei bedeuten:

E vor VF: Anzahl der Reihenfolgebeziehungen, in denen eine Kennzahl zur Erfolgslage vor einer Kennzahl zur Vermögens- und Finanzlage gebildet wird.

VF vor E: Anzahl der Reihenfolgebeziehungen, in denen eine Kennzahl zur Vermögens- und Finanzlage vor einer Kennzahl zur Erfolgslage gebildet wird.

Tabelle 4.1. zeigt das Ergebnis unserer Auswertung, gegliedert nach Menge, Inhalt und Reihenfolge der Kennzahlenvorschläge.

Tabelle 4.1.: Menge, Inhalt und Reihenfolge der Kennzahlenvorschläge bilanzanalytischer Monographien

Meßwerte	Minimum	25%-Perzentil	Median-Wert	75%-Perzentil	Maximum
Anzahl Kennzahlen	16	29	45	63	144
Anteil Kennzahlen zur Erfolgslage	20,6	35,5	41,5	45,7	61,5
Anzahl Kennzahlen Erfolgslage unter den ersten 10	0	0	2	6	10
Wechsel zwischen Analysefeldern	5,8	11,2	16,0	33,2	23,0
Indexwert Objektgliederung	-79,4	-55,6	-21,3	33,7	96,7

Zu a) Menge der Kennzahlenvorschläge

Die Spannweite der Kennzahlenmenge reicht von 16 bzw. 24 bei den Arbeitskreisen des Bundeswirtschaftsministeriums und des Rationalisierungskuratoriums der Wirtschaft bis zu 103 bzw. 144 bei den Lehrbüchern von Gräfer und Braune/Streck. Bemerkenswert ist, daß sich die Arbeitskreise an Praktiker wenden, die offenbar beachtliche Vorbehalte gegen "zu viele" Kennzahlen haben,⁶⁵⁾ während die anderen Autoren Vorschläge für eine computergestützte Bilanzanalyse unterbreiten, bei der das Errechnen von Kennzahlen keine Probleme aufwirft. Doch: Was helfen derartige "Kenn"-Zahlenmengen, wenn sie nicht genutzt werden können, weil sie die begrenzte Aufnahmefähigkeit des Analytikers überschreiten?

Zu b) Inhalt der Kennzahlenvorschläge

Auch hier gibt es große Unterschiede: Der Anteil von Kennzahlen zur Erfolgslage reicht von 20,6 bis 61,5 v. H. Er liegt jedoch bei den meisten Autoren zwischen 35 und 45 v. H. Die Bilanzanalyse wird offenbar primär als eine Vermögens- und Finanzanalyse begriffen, denn auf diese Analysefelder richten sich zwei Drittel der Kennzahlen.

Zu c) Reihenfolge der Kennzahlenvorschläge

Beurteilt man die ersten 10 Kennzahlenvorschläge eines Lehrbuchs nach einer 70:30-Regel, dann beginnen nur 6 Autoren mit der Erfolgslage, aber immerhin 16 mit der Vermögens- und Finanzlage. 2 lassen sich nicht eindeutig zuordnen. Der geringe Prozentsatz von Wechseln zwischen den Urteilsbereichen und die negativen Indexwerte der Skala "Erfolgslage vor Vermögens- und Finanzlage" lassen erkennen, daß diese Gliederung auch im Verlauf der weiteren Analyse eingehalten wird. Es ist jedoch darauf hinzuweisen, daß mehr Wechsel auftreten als bei den empirischen Untersuchungen des Informationsnachfrageverhaltens von Knorr und Weigel.⁶⁶⁾ Dies deutet darauf hin, daß sich die Kennzahlenbildung an Analysefragen orientiert und stärker von der Gliederung nach Rechenwerken löst als die Informationsbeschaffung, bei der eine angebotsgeleitete Vorgehensweise Rationalisierungsvorteile bietet. Der Befund bekräftigt zugleich unsere Bedenken, das menschliche Informationsverhalten allein von der Beschaffung her zu erfassen, wie dies in zahlreichen Prozeßstudien geschieht.

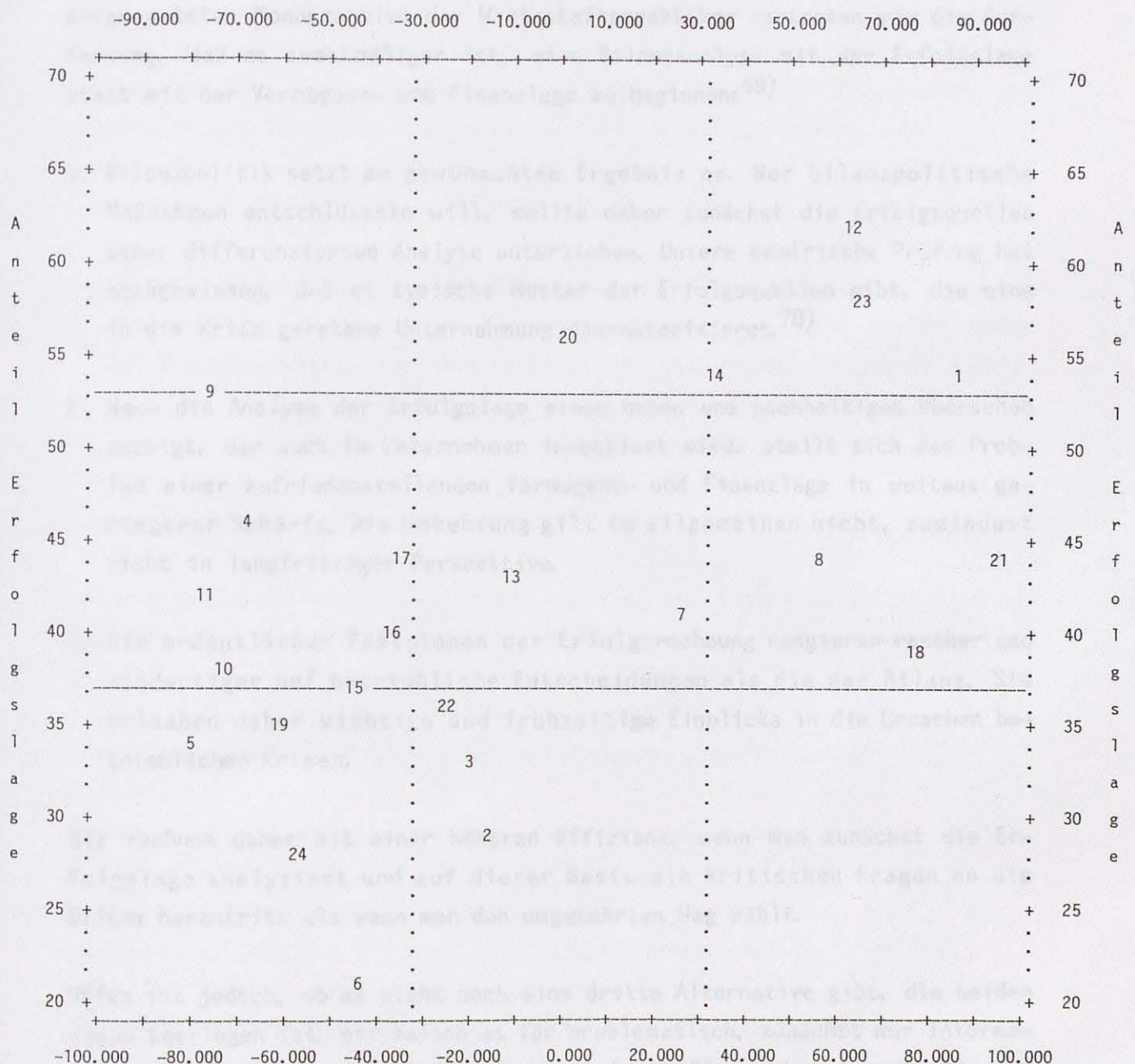
Auch der Befund zur Reihenfolge der Kennzahlenbildung zeigt erhebliche Unterschiede zwischen den Monographien. So streut der Index zur Objektgliederung zwischen - 79,6 und + 96,7. Für die "Mehrheitsmeinung", die sich vom 25-%-Perzentil bis zum 75-%-Perzentil erstreckt, ergibt sich ebenfalls eine recht beachtliche Bandbreite: Die Indexwerte reichen von -55,6 bis + 33,7. **Es gibt demnach keine einheitliche Meinung über das "richtige" Vorgehen bei der Bilanzanalyse.**

Um den unterschiedlichen Haltungen gerecht zu werden, versuchen wir unterschiedliche Typen zu identifizieren. Wir tragen die Monographien auf einem Koordinatenkreuz mit den Achsen "Anteil Kennzahlen zur Erfolgslage" und "Index der Objektgliederung" ein (vgl. Abbildung 4.3.).

Das Streupunktdiagramm läßt einen von links unten nach rechts oben verlaufenden Punkteschwarm erkennen, der links unten eine größere, recht kompakte und rechts oben eine kleinere, weniger deutlich ausgeprägt Häufung aufweist:

- * Zum oberen, **erfolgsorientierten Cluster** gehören Monographien, die **"aus der Praxis für die Praxis"** geschrieben wurden: namentlich die Monographie des ZVEI, die Schrift des Arbeitskreises beim Bundesminister für Wirtschaft und das Werk von Riebell für die Bilanzauswertung der Sparkassen (Nr. 23, 1, 18). Daß die Monographie von Lachnit (Nr. 12) zu diesem Cluster gehört, liegt an seinem tief gegliederten Kennzahlensystem zur Analyse der Erfolgslage, das an den in der Praxis verbreiteten Systemen anknüpft.⁶⁷⁾
- * Zum unteren, **vermögens- und finanzorientierten Cluster** gehören (weit verbreitete) **Lehrbücher** zur Bilanzanalyse, namentlich die Werke von Beyrer, Braune/Streck, Buchner, Coenenberg und Mitarbeitern, Erhard, Henseler, Hesse, Leffson, Ölschläger/Lodwitz, Perridon/Steiner, Reinheimer, Riemer, Schult und Zumbühl, sowie die Monographie von Hofmann.⁶⁸⁾ Auch wenn sich diese graduell unterscheiden, behandeln sie doch ausnahmslos die Vermögens- und Finanzlage vor der Erfolgslage. Mit Ausnahme von Henseler schlagen sie auch (wesentlich) mehr Kennzahlen zur Vermögens- und Finanzlage vor als zur Erfolgslage. Da sie dabei überwiegend auf das besonders heftig kritisierte Rechenwerk Bilanz (i.e.S.) zurückgreifen, kann man mit Fug und Recht feststellen, daß die deutschsprachigen Lehrbücher zur Bilanzanalyse im engeren Sinne "bilanzorientiert" sind. Wir können daher feststellen: **Wer "schul"-mäßig analysieren will, der sollte nach herrschender Meinung der bilanzanalytischen Lehrbücher zuerst und besonders intensiv Kennzahlen zur horizontalen und vertikalen Bilanzstruktur bilden, um die wirtschaftliche Lage eines Unternehmens zu beurteilen.**

Abbildung 4.3.: Inhalt und Reihenfolge der Kennzahlenvorschläge bilanzanalytischer Monographien



Legende:

1: Arbeitskreis "Kreditwürdigkeit durch Ertragskraft" beim Bundesminister für Wirtschaft 2: Beyer
 3: Braune/Streck 4: Buchner 5: Coenenberg/Brandi/Eifler/Schmidt 6: Erhard 7: Gräfer 8: Hecker
 9: Henseler 10: Hesse 11: Hofmann 12: Lachnit 13: Leffson 14: Lohse 15: Ölschläger/Lodwitz
 16: Perridon/Steiner 17: Reinheimer 18: Riebell 19: Riemer 20: Arbeitskreis "Mittel- und Kleinbetriebe" im Bundesausschuß Betriebswirtschaft des RKW 21: Schöning 22: Schult 23: Betriebswirtschaftlicher Ausschuß des Zentralverbandes der Elektrotechnischen Industrie e. V. 24: Zumbühl

Wir halten diese "Schulweisheit" für problematisch. Mit Hauschildt und den ausgewerteten Monographien der Wirtschaftspraktiker vertreten wir die Auffassung, daß es zweckmäßiger ist, eine Bilanzanalyse mit der Erfolgslage statt mit der Vermögens- und Finanzlage zu beginnen.⁶⁹⁾

1. Bilanzpolitik setzt am gewünschten Ergebnis an. Wer bilanzpolitische Maßnahmen entschlüsseln will, sollte daher zunächst die Erfolgsquellen einer differenzierten Analyse unterziehen. Unsere empirische Prüfung hat nachgewiesen, daß es typische Muster der Erfolgsquellen gibt, die eine in die Krise geratene Unternehmung charakterisieren.⁷⁰⁾
2. Wenn die Analyse der Erfolgslage einen hohen und nachhaltigen Überschuß anzeigt, der auch im Unternehmen investiert wird, stellt sich das Problem einer zufriedenstellenden Vermögens- und Finanzlage in weitaus geringerer Schärfe. Die Umkehrung gilt im allgemeinen nicht, zumindest nicht in langfristiger Perspektive.
3. Die ordentlichen Positionen der Erfolgsrechnung reagieren rascher und eindeutiger auf betriebliche Entscheidungen als die der Bilanz. Sie erlauben daher wichtige und frühzeitige Einblicke in die Ursachen betrieblicher Krisen.

Wir rechnen daher mit einer höheren Effizienz, wenn man zunächst die Erfolgslage analysiert und auf dieser Basis mit kritischen Fragen an die Bilanz herantritt als wenn man den umgekehrten Weg wählt.

Offen ist jedoch, ob es nicht noch eine dritte Alternative gibt, die beiden Wegen überlegen ist. Wir halten es für problematisch, zunächst **nur** Informationen der Gewinn- und Verlustrechnung zu beschaffen und zu verarbeiten und dann nur noch Informationen aus der Bilanz i.e.S. Wichtige **Beziehungszahlen**, namentlich Renditekennzahlen, Umschlagskennziffern und auf dem Cash-Flow aufbauende Verhältniszahlen, zeichnen sich gerade dadurch aus, daß sie Informationen aus beiden Rechenwerken verknüpfen. Dies legt den Gedanken nahe, daß eine **Vernetzung** beider Analyseobjekte erforderlich ist. So ist es z. B. sinnvoll, die Analyse **thematisch** zu ordnen und bei einem starken

Umsatzrückgang **sofort** nachzufragen, wie sich Reichweite des Erzeugnislagers und die Debitorenbestände entwickelt haben, um den Absatzbereich als Ganzes richtig einzuschätzen. In ähnlicher Weise kann man bei rückläufigem Betriebserfolg und sinkendem Personalanteil nach den Pensionsrückstellungen fragen, Veränderungen des restlichen Aufwandes können zur Entwicklung der anderen Rückstellungen in Beziehung gesetzt werden und vieles mehr.

Wir vertreten daher folgende Hypothese:

Wenn die Erfolgslage vor der Vermögens- und Finanzlage analysiert wird, dann ergibt sich eine höhere Effizienz der Analyse als bei der umgekehrten Vorgehensweise. Die höchste Effizienz wird jedoch bei einem mittleren Grad der Objektgliederung vermutet.

4.1.6. Die Einflüsse der Rahmenvariablen

In unseren Experimenten werden zahlreiche Einflüsse neutralisiert (vgl. Kapitel 4.2.). Es verbleiben zwei Größen, die von besonderem Interesse sind:

- * die experimentell variierte **Präsentationsform** der Daten und
- * die unterschiedliche **Erfahrung** der Versuchspersonen.

Damit wird je eine Variable der in der Literatur besonders häufig untersuchten Einflußblöcke **"Person"** und **"Informationssystem"** berücksichtigt.

Anzumerken ist, daß die Versuchspersonen unterschiedliche **Fälle** bearbeiten, so daß auch ein Merkmal des Einflußblocks **"Aufgabe"** variiert wird. Zum Falleinfluß werden keine Hypothesen abgeleitet. Unser Effizienzmaß ist so konstruiert, daß im Durchschnitt aller Versuchspersonen keine Falleinflüsse zu erwarten sind. (Vgl. hierzu die Modellrechnungen in Kapitel 3.2.5.1.). Daß die Effizienzmeßwerte über die experimentellen Fälle hinweg zunehmen, kann als **Lernfortschritt** interpretiert werden. (Vgl. hierzu die Befunde in Kapitel 3.4.). Hierzu werden keine neuen Hypothesen abgeleitet. Wir diffe-

4.1. Theoretischer Bezugsrahmen der empirischen Untersuchung Seite - 319 -
renzieren vielmehr zwischen einer interpersonellen und einer intertemporalen Interpretation der bereits abgeleiteten Hypothesen und prüfen diese Varianten mit unterschiedlichen Testdesigns (vgl. Kapitel 4.3.). Außerdem wird geprüft, ob es systematische Wechselwirkungen zwischen dem Falleinfluß und den Einflüssen des Informationsverhaltens auf die Effizienz gibt.

4.1.6.1. Die Einflüsse der Präsentationsform

In unseren Experimenten werden den Versuchsgruppen die aufbereiteten Daten in einer **Baumstruktur** angeboten. Den Kontrollgruppen werden die inhaltlich und numerisch gleichen Daten in einer tabellarischen Form präsentiert (vgl. Kapitel 4.2.). Unsere Forschungsfrage lautet daher: Welche Unterschiede im Informationsverhalten und in der Effizienz sind zu erwarten, wenn Kennzahlen in einer Baumstruktur statt in einer tabellarischen Darstellung präsentiert werden?⁷¹⁾

Bei der Beantwortung dieser Frage können wir auf die Analyse von **Hauschildt** zurückgreifen, der die Einflüsse auf die Effizienz systematisch überprüft hat.⁷²⁾ Nicht behandelt werden die Einflüsse der Baumstruktur auf die Verrichtungs- und Objektgliederung des Informationsprozesses, da diese Gegenstand der Arbeit von Petersen sind.

a) Einflüsse auf die Effizienz der Analyse

Wirtschaftspraktiker und betriebswirtschaftliche Forscher gehen schon seit langem davon aus, daß baumstrukturartige Informationssysteme die Effizienz von Beurteilungsprozessen steigern, indem sie sowohl die **Differenzierung** als auch die **Integration** fördern.⁷³⁾

- * In disaggregierender Richtung regen Baumstrukturen an, unwichtige Teilinformationen auszusondern und wichtige zu differenzieren.
- * In aggregierender Richtung fördern sie die Verknüpfung und Relativierung von Teilurteilen und helfen abzuschätzen, wie stark spezifische Ursachen auf das Gesamtergebnis durchschlagen.

Diese beiden Vorgänge sind nicht isoliert zu sehen. Sie sollten auch nicht in einer strengen Phasenfolge ablaufen. Es kommt vielmehr auf das praktisch gleichzeitige Auf und Ab an, auf die Differenzierung **und** Integration an. Erst dies bewirkt eine nachhaltige Verbesserung des Beurteilungsprozesses.

Nach Hauschildt sind tabellarische Darstellungen bei der Unterstützung dieses Vorgangs unterlegen, weil die Tabelle weniger effektive, **zusätzliche "Hilfs"-Informationen** benötigt, um die **Verknüpfung** der Informationen sichtbar zu machen.⁷⁴⁾

Die weite Verbreitung von Baumstrukturen bei der Suche von Fehlern in der Hard- und Software komplexer technischer Systeme kann demgegenüber vor allem auf eine Beschleunigung des Suchprozesses und eine Verringerung seiner Kosten zurückgeführt werden.⁷⁵⁾

Die Vorteile einer Baumstruktur sind nicht unumstritten. So belegen Fischhoff, Slovic und Lichtenstein, die mehrere Experimente mit unterschiedlich breiten und unterschiedlich tiefen "fault trees" durchführen, folgende dysfunktionale Effekte:⁷⁶⁾

1. Variation der Baumbreite (bei mittlerer Tiefe):

Fehlerursachen, die nicht in der präsentierten Baumstruktur enthalten sind, werden stark vernachlässigt. In unvollständigen Bäumen präsentierte Fehlerursachen werden als wesentlich wahrscheinlicher angesehen als in vollständig präsentierten Bäumen. Dies gilt auch für Experten.⁷⁷⁾ Wenn man die Versuchspersonen explizit darauf hinweist, daß die Bäume unvollständig seien, nimmt der Präsentationseffekt zwar deutlich ab, aber er ist immer noch sehr hoch.

2. Variation der Baumtiefe (bei voller Breite):

Die Variation der Tiefe des gesamten Baums oder ausgewählter Äste zeigt schwache Effekte auf die wahrgenommene Wahrscheinlichkeit der (jeweiligen) Fehlerursachen. Als konsistenter Effekt kann nur festgestellt werden, daß

4.1. Theoretischer Bezugsrahmen der empirischen Untersuchung Seite - 321 -
die (ohnehin recht geringe) Wahrscheinlichkeit der Restkategorie ("other") mit steigender Tiefe abnimmt.

3. Variation der Baumstruktur: Splitten und Zusammenfassen von Ästen

Wenn Äste gesplittet werden, erhöht sich die wahrgenommene Fehlerwahrscheinlichkeit, bei Zusammenfassung verringert sie sich – beide Änderungen verglichen mit dem Ausgangsbaum. Der Einfluß ist schwächer als eine Variation der Baum-"Breite" und stärker als eine Variation der Baum-"Tiefe".

Aus den Ergebnissen dieser Experimentalserie kann man nicht schließen, daß eine tabellarische Struktur effizienter sei als eine Baumstruktur – denn diese Gegenüberstellung wird nicht getestet – aber man ist gewarnt, daß die graphische Suggestionskraft von Baumstrukturen auch dazu führen kann, daß Probleme selektiv und unvollständig wahrgenommen werden.

In unseren Experimenten stellt sich die Frage nach der Vollständigkeit des Baums in einer anderen Weise als bei Fischhoff/Slovic/Lichtenstein: Das von uns verwendete mathematisch strukturierte Kennzahlensystem ist im Unterschied zu einem Ordnungssystem definitions-logisch vollständig, und zwar in der Breite.⁷⁸⁾ Inwieweit der Analytiker bei seiner Analyse allerdings zu materiellen und bilanzpolitischen Ursachen vordringt, dürfte auch eine Frage der Baumtiefe und der Art der Baumstruktur sein. Da die Baumtiefe nach Fischhoff/Slovic/Lichtenstein von geringerer Bedeutung ist, kommt es offenbar in unserem Fall besonders auf die Qualität der Baumgliederung an. Wir können hierzu auf unsere empirische Validierung der Baumstruktur verweisen, die auch den Befunden anderer Forscher entspricht.⁷⁹⁾

Wir stellen daher folgende Hypothese auf:

Die Präsentation eines Informationsangebotes in Baumstrukturen führt zu einer höheren Effizienz der Analyse als die Präsentation in Tabellen.

b) Einflüsse auf das Informationsverhalten

Eine Baumstruktur bietet dem Analytiker ein extern vorgedachtes Modell an, mit dem er seinen Analyseprozeß sinnvoll nach verschiedenen Modulen gliedern kann. Als Suchhilfe erlaubt sie ihm, irrelevante Äste auszuschließen und schneller zu den Ursachen vorzudringen.⁸⁰⁾ Als Integrationshilfe regt sie an, die auf unterschiedlichen Aggregationsebenen getroffenen Aussagen miteinander zu verknüpfen.⁸¹⁾

Wir rechnen mit einer selektiven, zielgerichteten Informationsbeschaffung und Übertragung und einem höheren Niveau der Informationsverarbeitung: Es werden weniger Informationen beschafft und übertragen, aber diese werden durch mehr Kennzahlen und Beurteilungen miteinander verknüpft. Die freigesetzte Kapazität kann genutzt werden, komplexere Vergleichstechniken anzuwenden und stärker in die Tiefe zu gehen, d. h. mehr unverdichtete Informationen zu analysieren. Dies entspricht folgenden Hypothesen:

Die Präsentation eines Informationsangebotes in Baumstrukturen führt zu

- a) einer geringeren Anzahl beschaffter Informationen,
- b) einer geringeren Anzahl übertragener Informationen,
- c) einer größeren Anzahl gebildeter Kennzahlen,
- d) einer höheren Komplexität der Vergleichstechnik und
- e) einem geringeren Aggregationsgrad der Informationsaktivitäten

als die Präsentation in Tabellen.

4.1.6.2. Die Einflüsse der Erfahrung

a) Einflüsse auf die Effizienz

Analytiker mit einer größeren Erfahrung können die Informationen zielgerichteter beschaffen, übertragen und zu Kennzahlen verdichten. Sie werden daher weniger Daten bereitstellen, die sie nicht benötigen oder nicht mehr verarbeiten können. Sie dürften die Aufbereitung ihrer Informationspakete

schneller abschließen. Wegen dieser **Kapazitäts-Erklärung** besitzen erfahrene Analytiker ein höheres Potential, um zu effizienteren Analysen zu gelangen.

Eine gute Bilanzanalyse ist nicht nur eine Frage der Rationalisierung von Informationsaktivitäten. Es kommt auch darauf an, die gewonnenen Informationen zu **nutzen**. Ein erfahrener Analytiker dürfte über ein höheres **Faktenwissen** verfügen, an dem er konkrete Kennzahlenausprägungen relativieren kann. Wissen besteht jedoch nicht nur aus ontologischem Faktenwissen und normativen Bewertungsprämissen.⁸²⁾ Entscheidend ist unseres Erachtens das **Beziehungswissen**. Es enthält sowohl definitorische Teil-Ganzes-Relationen als auch kausale Hypothesen.⁸³⁾ Experten zeichnen sich dadurch aus, daß sie im Langzeitgedächtnis über differenziertere Netze von Informationsverknüpfungen verfügen. So schätzt Simon, daß ein Großmeister des Schachs ca. 50.000 Muster gespeichert hat.⁸⁴⁾ In den frühen "Expertensystemen" MYCIN und INTERNIST sollen nach seinen Angaben einige Hundert Beziehungen gespeichert sein.⁸⁵⁾ Ohne uns auf eine konkrete Zahl festlegen zu müssen, können wir die Hypothese aufstellen, daß differenzierte Gedächtnisstrukturen den Analytiker in die Lage versetzen, komplexere und spezifischere Analyseprozesse zu absolvieren. So stellt insbesondere Bouwman fest, daß erfahrene Bilanzanalytiker komplexere Verknüpfungen zwischen den angebotenen Bilanzdaten herstellen, tiefer nach Ursachen forschen und einmal aufgestellte Hypothesen an anderen Sachverhalten überprüfen. Naive Analytiker zeichnen sich demgegenüber durch wesentlich einfachere, lineare Denkabläufe aus.⁸⁶⁾ Diese Vermutung wird durch die Nachfolgestudie von Anderson gestützt.⁸⁷⁾ Ähnliche Ergebnisse zeigen Studien psychologischer Forscher.⁸⁸⁾

Wir vermuten daher, daß Analytiker mit einer größeren Erfahrung auch bei gleicher Menge und Qualität an beschafften, übertragenen und zu Kennzahlen verdichteten Informationen bessere Analysen anfertigen:⁸⁹⁾

Je größer die Erfahrung des Analytikers, desto größer die Effizienz der Analyse.

b) Einflüsse auf das Informationsverhalten

Analytiker mit größerem Vorwissen können reichhaltigere Verknüpfungen zwischen den Daten herzustellen. Wir rechnen daher mit einer komplexeren Vergleichstechnik und einem höheren Niveau der Kennzahlenbildung.

Bei der Informationsbeschaffung sind gegenläufige Effekte abzuwägen: Erfahrene Analytiker können gezielter nachfragen, benötigen also weniger Informationen. Dagegen fehlt unerfahrenen Analytikern das Vorwissen, um den sachlichen Gehalt und die Bedeutsamkeit bestimmter Kennzahlenausprägungen einschätzen zu können.⁹⁰⁾ Sie werden sich daher auf vertraute Positionen zurückziehen, deren betriebswirtschaftliche Bedeutung offenkundig ist. Welcher Effekt überwiegt, ist schwer abzuschätzen: Eine Bestandsaufnahme vorliegender Ergebnisse aus der empirischen Entscheidungsforschung zeigte keinen signifikanten Zusammenhang zwischen der Entscheidungshäufigkeit und der Menge an nachgefragten Informationen.⁹¹⁾ In unserem Experiment stellt jedes Unternehmen einen neuen Fall dar. Es entfällt der Rückgriff auf ein umfangreicheres, mental gespeichertes Faktenwissen bei erfahreneren Personen. Per Saldo rechnen wir mit einer geringeren Menge an beschafften und übertragenen Informationen und einem höheren Anteil verdichteter Daten bei unerfahrenen Analytikern. Dies entspricht folgenden Hypothesen:

Je größer die Erfahrung des Analytikers,

- a) desto mehr Informationen werden beschafft,**
- b) desto mehr Informationen werden übertragen,**
- c) desto mehr Kennzahlen werden gebildet,**
- d) desto höher die Komplexität der Vergleichstechnik und**
- e) desto geringer der Aggregationsgrad der Informationsaktivitäten.**

4.1.7. Zusammenfassung der Hypothesen

Damit ist unser theoretisches Modell vollständig formuliert. Die folgende Hypothesenmatrix faßt die abgeleiteten Beziehungen zusammen (Abbildung 4.5.).

Abbildung 4.4.: Zusammenfassende Darstellung der abgeleiteten Hypothesen

abhängige Variable	unabhängige Variable				
	Rahmenvariablen		Strategische Dimensionen		Taktische Dimensionen
	Präs. Erf.		Obj. Verr.	Vgl. Aggr.	Operative Dimensionen
Strategische Dimensionen Objekt-Gliederung Verrichtungs-Gliederung	P P	P P	P 0 0	0 0	
Taktische Dimensionen Vergleichstechnik Aggregationsgrad	+	+	0	0	+
Operative Dimensionen Beschaffung Übertragung Kennzahlenbildung	-	+	0	0	0
	-	+	0	0	0
	+	+	0	0	0
Effizienz der Analyse	+	+	n	n	n
					n
					n
					+ / n
					+ / n
					+ / n

Legende:

0: Kein systematischer Zusammenhang

+: Positiver Zusammenhang

n: umgekehrt u-förmiger Zusammenhang

P: Gegenstand der Arbeit von Petersen

4.2. METHODIK UND DURCHFÜHRUNG DER EMPIRISCHEN UNTERSUCHUNG

4.2.1. Überblick

Die Prüfung der soeben abgeleiteten Hypothesen erfordert eine Festlegung auf eine konkrete Aufgabe, bestimmte Versuchspersonen, eine bestimmte Erhebungssituation und bestimmte Erhebungs-, Skalierungs- und Auswertungsverfahren. Abbildung 4.5. zeigt die damit verbundenen Probleme und unsere jeweilige Wahl. In den nachfolgenden Kapiteln werden die gewählten Alternativen näher beschrieben und die getroffene Wahl begründet.

Mit den einzelnen Schritten sind unterschiedliche methodische Probleme verknüpft, die in einer für den Anwender kaum noch überschaubaren Spezialliteratur behandelt werden. Wir orientieren uns an den von Cook und Campbell aufgestellten Forderungen.¹⁾ Sie unterscheiden vier Arten von Validitätsbegriffen:²⁾

1. **Externe Validität ("external validity"):** Hierunter verstehen Cook und Campbell die Eigenschaft, eine kausale Beziehung über verschiedene Personen, Kontexte ("settings") und Zeitpunkte hinweg zu generalisieren. Für uns stellt sich die Frage, inwieweit man die empirisch festgestellten Zusammenhänge zwischen Informationsverhalten und Effizienz der Bilanzanalyse auf Wirtschaftspraktiker, ähnliche Beurteilungsprobleme und andere Informationsangebotssituationen übertragen kann. Diese Fragen werden in den Kapiteln 4.2.2. bis 4.2.4. untersucht.
2. **Interne Validität ("internal validity"):** Hierunter ist die Eigenschaft eines Versuchsaufbaus zu verstehen, gültige Variationen der untersuchten unabhängigen Variablen zu bewirken und systematische Verzerrungen nicht betrachteter Größen auszuschließen, um die (kausale) Eindeutigkeit der Interpretation der Ergebnisse im Rahmen des Versuchsaufbaus zu sichern. Fragen dieser Art werden im Gliederungsabschnitt 4.2.5. diskutiert.
3. **Konstruktvalidität ("construct validity"):** Sie bezeichnet die Eigenschaft einer Operationalisierung, theoretische Begriffe inhaltlich zutreffend, umfassend und trennscharf gegenüber Operationalisierungen an-

Abbildung 4.5.: Methodische Entscheidungsprobleme und gewählte Alternativen

Aktionsparameter des Designs	Methodisches Problem	Gewählte Alternative
Art der Aufgabe	Externe Validität	Individuelle Bilanzanalyse zur Vorbereitung einer Kreditentscheidung
Art der Subjekte	Externe Validität	Studentische Versuchspersonen der Wirtschaftswissenschaft (durchschnittlich 7,0 Fachsemester)
Erhebungssituation	Externe und Interne Validität	Standardisierte Laborexperimente mit einem realistischen Informationsangebot
Erhebungsverfahren	Reliabilität und Konstruktvalidität	Informationsbeschaffung: Information-Display-Boards; andere Informationsaktivitäten: Systematische Inhaltsanalyse der angefertigten Bilanzanalysen
Skalierungsverfahren	Konstruktvalidität	Eigens entwickelte Programme zur Bestimmung von Meßwerten der Intensität, Struktur und Reihenfolge des Informationsverhaltens
Auswertungsverfahren	Validität des statistischen Schlusses	Pfadanalysen, differenziert nach fallweisen, durchschnittlichen und ipsativen Meßwerten sowie nach Meßwerten zur Gesamt-, Erfolgs-, Vermögens- und Finanzlage

derer Begriffe zu messen. Für die Effizienzmessung wurde diese Anforderung im dritten Teil der Arbeit geprüft. Die Messung des Informationsverhaltens wird im Kapitel 4.2.6. behandelt.

4. **Validität des statistischen Schlusses ("statistical conclusion validity"):** Sie bezieht sich auf die Anwendung problemangemessener Verfahren und die Erfüllung von Prämissen statistischer Modelle. Validitätsprobleme dieser Art werden im Kapitel 4.2.7. gewürdigt.

4.2.2. Das Untersuchungsobjekt

Wie fast alle empirischen Untersuchungen beschränkt sich auch die vorliegende auf eine bestimmte Art von Aufgabe. Dies ist in unserem Fall eine individuell anzufertigende Bilanzanalyse zur Vorbereitung einer Kreditentscheidung. Die konkrete Aufgabenformulierung ist in Abbildung 4.6. dargestellt.

Unserer Entscheidung lagen folgende **Anforderungen** zugrunde:

1. **Die zu untersuchende Aufgabe sollte ein beachtliches betriebswirtschaftliches Problem darstellen.** Diese Anforderung schloß alle nicht-ökonomischen (technischen, medizinischen, juristischen, psychologischen, oder rein formal-logischen) Entscheidungsprobleme aus.
2. **Die zu untersuchende Aufgabe sollte eine beachtliche Informationsproblematik aufwerfen.** Aus einem umfangreichen Informationsangebot sollten Informationen unterschiedlicher Qualität aktiv gesucht und selbständig verknüpft werden. Diese Anforderung schloß die weit verbreiteten "Linsen-Modelle" aus, bei denen eine kleine Menge vorgegebener Kennzahlen zu einem Gesamturteil zu integrieren ist.
3. **Die Effizienz der zu untersuchenden Aufgabe sollte vornehmlich vom individuellen Informationsverhalten beeinflusst werden.** Diese Anforderung schloß alle Aufgaben aus, bei denen Durchsetzungsprobleme und Fragen der Willensbildung im Vordergrund stehen.

Abbildung 4.6.: Die Aufgabenstellung

AUFGABENSTELLUNG

Stellen Sie sich vor, Sie seien Kreditsachbearbeiter in einer Bank. Sie finden am Morgen, als Sie Ihren Dienst antreten, zwei Notizen auf Ihrem Tisch vor:

1. Notiz Ihres Vorgesetzten

"Ich habe gestern mit Herren der ... AG zu Mittag gegessen und anschließend das Werk in T-hausen besichtigt. Ich hatte einen ausgezeichneten Eindruck von der Firma. Das Unternehmen ist straff und zielstrebig geführt. Die Herren werden heute ihre Bilanzen einreichen und sehen der Kreditverhandlung in der nächsten Woche entgegen."

2. Notiz eines Kollegen aus der Filiale Ihrer Bank in T-hausen

"Unter Ihren Kunden ist doch die ... AG. Gestern hörte ich vertraulich, daß es dieser Unternehmung nicht gut gehen soll. Es fiel der Satz: 'Die Ratten verlassen das sinkende Schiff!' Ich halte es für nützlich, daß Sie sich die Daten dieser Unternehmung einmal kritisch ansehen."

Am Mittag desselben Tages liegen Ihnen die Jahresabschlüsse der Jahre .. und ... der ... AG sowie die Daten der Branche von ... vor.

* Beurteilen Sie die Unternehmung, und begründen Sie Ihr Urteil.

* Formulieren Sie eine Liste offener Fragen, die Sie aus den Bilanzmaterialien nicht beantwortet erhalten, damit Ihr Chef in der folgenden Kreditverhandlung mit der ... AG die Antworten direkt einholen kann.

4. Für die Aufgabe sollten Effizienzmaßstäbe vorliegen bzw. erarbeitet werden können. Diese Anforderung erschien uns unverzichtbar, weil wir Effizienzhypothesen aufstellen und überprüfen wollten, um daraus Gestaltungshinweise abzuleiten.

5. Die zu untersuchende Aufgabe sollte unter kontrollierten, reproduzierbaren Bedingungen studiert werden können. Diese Anforderung wurde aufgestellt, um im Rahmen mehrerer Versuchsserien mit vergleichbarem Grunddesign die Wirkungen verschiedener Rahmenvariablen systematisch erforschen zu können.

6. Das Informationsverhalten während der Aufgabenlösung sollte prozeßbegleitend und differenziert gemessen werden können. Diese Forderung schloß die wesentlich schwieriger zu erfassenden multipersonalen Probleme aus. Wir entschieden uns vielmehr für individuelle Entscheidungsprobleme.³⁾

7. Die zu untersuchende Aufgabe sollte bei verschiedenen Fällen mit formal gleichen Anforderungen, aber inhaltlich unterschiedlich ausgeprägten Informationen studiert werden können. Diese Anforderung wurde aufgestellt, weil mehrere Studien gezeigt hatten, daß interindividuelle Leistungsunterschiede bei formal gleichen, aber inhaltlich verschiedenen Aufgaben sehr unterschiedlich ausfallen können.

8. Die zu untersuchende Aufgabe sollte sowohl durch Schulungsmaßnahmen als auch durch Informationshilfen unterstützt werden können. Diese Anforderung wurde aufgestellt, weil wir differenziert überprüfen wollten, wie zwei häufig idealtypisch unterschiedene Strategien "Veränderung des Menschen" und "Veränderung der Informationsbedingungen" auf das tatsächliche Informationsverhalten und dessen Effizienz wirken.⁴⁾

Eine Aufgabe, die diese Anforderungen erfüllt, stellt die Bilanzanalyse dar:

1. Bilanzanalysen sind substantielle betriebswirtschaftliche Probleme. Sie werden vielen wichtigen Entscheidungen der Wirtschaftspraxis zugrundegelegt und im betriebswirtschaftlichen Schrifttum ausführlich gewürdigt.

2. Die Bilanzanalyse wirft eine beachtliche Informationsproblematik auf. Bei der Bilanzanalyse ist eine Vielzahl von Informationen zu verdichten und miteinander zu verknüpfen. Um die gebildeten Kennzahlen zu **beurteilen** und die spezifischen Stärken und Schwächen des Einzelfalls herauszuarbeiten, benötigt man viel Erfahrung und teilweise recht komplexe Vergleichsketten. Die qualitativ heterogenen Bilanzdaten verlangen eine sorgfältige Abschichtung der Aussagen.

3. Die Qualität der Bilanzanalyse hängt vom Informationsverhalten im Analyseprozeß ab. In der betriebswirtschaftlichen Fachliteratur finden sich zahlreiche Aussagen, in denen Effizienzbehauptungen über die bei der Bilanzanalyse auszuführenden Informationsaktivitäten ausgesprochen werden. Sie reichen von der Beschaffung, Auswahl und Bereinigung des Urmaterials bis zur Bildung und Gewichtung der zu verwendenden Kennzahlen. Es fehlen jedoch Studien, in denen die meist implizit aufgestellten Effizienzbehauptungen **empirisch überprüft** werden.⁵⁾

4. Für die Bilanzanalyse können zuverlässige und valide Effizienzkriterien bestimmt werden. Die Erfüllung dieser Anforderung erwies sich als unerwartet schwierig: Wenn über die Zwecke von Bilanzanalysen gesprochen wird, dann werden entweder nur sehr allgemeine Überlegungen angestellt, aus denen man noch keine operationalen Effizienzmaßstäbe ableiten kann, oder ein Teilprozeß der Analyse, namentlich das Berechnen und Bewerten von Kennzahlen, wird zum Selbstzweck erhoben. Eine dritte Auffassung weist Bilanzanalysen die Aufgabe zu, ein "richtiges", klassifikatorisches Urteil zu begründen. Wir meinen, daß Bilanzanalysen ungeeignet sind, definitive Urteile abzugeben. Sie sollen vielmehr Fragen aufwerfen, weitere Informationsaktivitäten steuern und differenzierte Hinweise auf mögliche Maßnahmen geben. Es ist daher erforderlich gewesen, ein eigenes Meßkonzept zu entwickeln, das die **Qualität der Entscheidungsvorbereitung** mißt. Wir verweisen auf den dritten Teil unserer Arbeit.

5. Die Bilanzanalyse kann unter kontrollierten, reproduzierbaren experimentellen Bedingungen studiert werden. Auch die Erfüllung dieser Anforderung war schwierig: Die Entwicklung eines Informationslabors, das reproduzierbare Experimente ermöglicht, führte zu mehrjährigen Testversuchen und Pretests, bevor man die hier ausgewerteten Experimente durchführen konnte.

6. Die Bilanzanalyse eignet sich für eine prozeßbegleitende und differenzierte Messung des Informationsverhaltens. Sie ist einerseits hinreichend komplex, um ein anspruchsvolles Informationsproblem aufzuwerfen, andererseits aber noch so überschaubar, daß sie von einem einzelnen Analytiker in einem begrenzten Zeitraum bearbeitet werden kann. Wir können die Informationsbeschaffung mittels sogenannter "Information-Display-Boards" messen. Die anderen Informationsaktivitäten werden durch eine systematische Inhaltsanalyse der Analysebögen erhoben.

7. Die Bilanzanalyse eignet sich für einen Vergleich von Problemen, die formal gleiche Anforderungen stellen, aber inhaltlich unterschiedlich ausgeprägt sind. Bei der Bilanzanalyse kann man mehrere Fälle mit unterschiedlich ausgeprägten Informationen analysieren lassen. Auf diese Weise kann man die Konsistenz interindividueller Unterschiede feststellen.

8. Die Bilanzanalyse kann sowohl durch Schulungsmaßnahmen als auch durch Informationshilfen wirksam unterstützt werden. Bei der Bilanzanalyse sind zahlreiche einfache arithmetische und logische Operationen erforderlich, die an Computer delegiert werden können. Von dieser Datenverdichtungsmöglichkeit wird in den Experimenten Gebrauch gemacht. Sie beschränkt sich auf Summen und Salden, weil die Versuchspersonen Verhältniszahlen (Indexzahlen, Gliederungszahlen, Beziehungszahlen) selbst bilden sollen, um ihre Informationsbedürfnisse und ihren bilanzanalytischen Sachverstand erforschen zu können. Als graphische Informationshilfe werden die experimentell variierten Präsentationsformen "Baum" und "Tabelle" eingesetzt. Um festzustellen, wie Schulungsmaßnahmen wirken, wird nach jeder Analyse eine 90-minütige bilanzanalytische Feedback-Veranstaltung durchgeführt.

Fazit: Die Bilanzanalyse erfüllt die gestellten Anforderungen gut bis sehr gut. Die unerwartet umfangreichen Vorarbeiten sind ein sinnvoller Preis dafür, daß wir ein **reales** und nicht ein künstliches, experimentell vereinfachtes Problem untersuchen.

Man kann das entwickelte Instrumentarium bei analogen Problemen des Controlling, der Revision, der Marktforschung und der strategischen Frühaufklärung einsetzen. Auf der anderen Seite ist auf Grenzen der Generalisierbarkeit der empirischen Prüfung hinzuweisen. Eine allgemeingültige Aussage läßt sich nicht machen, denn empirisch geprüfte Taxonomien von Problemen liegen erst in Ansätzen vor.⁶⁾ Wir beschränken uns daher auf eine tabellarische Übersicht, die nicht nur die Grenzen unserer Untersuchung aufzeigt, sondern auch illustriert, wieviele Faktoren man konstant halten muß, um eindeutige Schlüsse ableiten zu können (vgl. Abbildung 4.7.).⁷⁾

4.2.3. Die Versuchspersonen

Unsere Versuchspersonen haben mehrheitlich 5 bis 8 Semester Betriebswirtschaftslehre mit Spezialisierung im Fach Unternehmensrechnung studiert, den Eingangstest über Inhalt und Bewertung von Jahresabschlüssen bestanden und sind als überdurchschnittlich test-intelligent anzusehen.⁸⁾

Im einzelnen haben an den Veranstaltungen folgende Versuchspersonen teilgenommen:

- * 34 Studenten der Wirtschaftswissenschaft der Universität Gießen im Wintersemester 1981/82 (durchschnittlich: 7,5 Fachsemester). Ihre Bilanzanalysen werden unter der Bezeichnung **Experimentalserie Gießen 82** zusammengefaßt.
- * 33 Studenten der Wirtschaftswissenschaft der Universität Kiel im Sommersemester 1982 (durchschnittlich: 6,5 Fachsemester). Ihre Bilanzanalysen werden unter der Bezeichnung **Experimentalserie Kiel 82** zusammengefaßt.

Abbildung 4.7.: Grenzen der Generalisierbarkeit auf andere Probleme

Merkmal	Konsequenz für die Generalisierbarkeit
Individuelle Entscheidung	Keine Generalisierung auf soziale Interaktions- und Kommunikationsprozesse
Beurteilungsproblem	<ul style="list-style-type: none"> a) "Extensive", kognitiv gesteuerte Problemlösung: keine Generalisierung auf "habituelle" oder "impulsive" Entscheidungen b) Keine Generalisierung auf Auswahlprobleme, da keine Auswahl zwischen (gleichzeitig vorliegenden, ohne Rückgriffsrestriktion verfügbaren) Handlungs-Alternativen
Passives, rein numerisches, gebührenfreies, visuelles Informationsangebot	<ul style="list-style-type: none"> a) Keine Generalisierung auf Mensch-Maschine-Interaktionen b) Keine Generalisierung auf Entscheidungen, bei denen Informations-Quellen aktiv gesucht und bewertet werden c) Keine Generalisierung auf Informations-Entscheidungssituationen, bei denen die Informationsbeschaffung (erhebliche) Kosten verursacht d) Keine Generalisierung auf Informationssituationen, in denen bildhafte oder nicht-visuelle Reize präsentiert und aufgenommen werden
Ohne Zeitdruck	Keine Generalisierung auf Entscheidungen, in denen der Zeitdruck eine erhebliche Rolle spielt

* 36 Studenten der Wirtschaftswissenschaft der Universität Kiel im Sommersemester 1984 (durchschnittlich: 6,4 Fachsemester). Ihre Bilanzanalysen werden unter der Bezeichnung **Experimentalserie Kiel 84** zusammengefaßt.

Über die prinzipielle Eignung der Versuchspersonen, den Zusammenhang zwischen Informationsverhalten und Effizienz anhand des komplexen betriebswirtschaftlichen Beurteilungs- und Diagnoseproblems "Bilanzanalyse" zu untersuchen, dürfte kein Zweifel bestehen. Offen ist der stets unsichere Induktionsschluß auf "Praktiker",⁹⁾ die Bilanzanalysen zur Entscheidungsvorbereitung verwenden. Campbell und Stanley schreiben mit wünschenswerter Klarheit:

"**Logically**, we cannot generalize beyond these limits; i. e. we cannot generalize at all."¹⁰⁾

Bei diesem Induktionsschluß muß man auf Plausibilitätsannahmen und abgeschätzte Gesetzmäßigkeiten verlassen.¹¹⁾ Dies wollen wir nicht tun. Uns kommt es darauf an, daß man überhaupt erst einmal den Zusammenhang von Informationsverhalten und Effizienz in einer Versuchsanordnung, die dem Sachverhalt gerecht wird, empirisch prüft. Primäres Ziel ist der **Test eines komplexen Hypothesensystems**. Erst wenn dieses System nicht falsifiziert wird, sollte man Experimente mit Praktikern, die sehr schwer für Laborexperimente zu gewinnen sind, durchführen. Es wäre unzweckmäßig, den zweiten vor dem ersten Schritt zu planen. Trotzdem wollen wir uns der Diskussion um die Populationsvalidität nicht entziehen. Wir behandeln zunächst die Vielschichtigkeit dieser Problematik, dann die empirische Evidenz und schließlich forschungspragmatische Aspekte.

a) Vielschichtigkeit der Generalisierungsproblematik¹²⁾

Man darf davon ausgehen, daß Wirtschaftspraktiker, die ihren Entscheidungen üblicherweise Bilanzanalysen zugrundelegen, im Durchschnitt bessere Analysen anfertigen als Studenten. Dies heißt aber nicht, daß jeder Wirtschaftspraktiker eine bessere Bilanzanalyse anfertigt als jeder einigermaßen fachlich vorgebildete Student der Betriebswirtschaftslehre. Es gibt

zwischen beiden Gruppen Überschneidungen und innerhalb beider Gruppen erhebliche Unterschiede. Daher ist es auch innerhalb **beider** Gruppen möglich, Zusammenhänge zwischen Informationsverhalten und Effizienz zu prüfen.

Das Generalisierungsproblem besteht **nicht** darin, daß Praktiker bessere Analysen anfertigen (können). Unser Hypothesensystem behauptet ja ausdrücklich, daß erfahrene Analytiker bessere Analysen anfertigen werden, wenn sie gleich viele Informationen beschaffen, übertragen und zu Kennzahlen verdichten. Das Generalisierungsproblem besteht vielmehr darin, daß wir nicht wissen, ob sich bei Konstanzhaltung der Erfahrung bei Wirtschaftspraktikern ein anderer **Zusammenhang** zwischen Informationsverhalten und Effizienz ergeben wird als bei den hier untersuchten studentischen Analytikern. In der Sprache der Varianzanalyse handelt es sich um einen **Interaktionseffekt**.¹³⁾

Nun ist es **nicht** damit getan, die gleichen Experimente mit Wirtschaftspraktikern durchzuführen. Dies würde nur **einen** spezifischen Interaktionseffekt testen, nämlich den der Populationszugehörigkeit in einer standardisierten Laborsituation. Es verbleiben die Generalisierungsprobleme, daß "die" Praktiker in ihren unterschiedlichen Organisationen von unterschiedlichen Motivations- und Sanktionssystemen beeinflusst werden, Bilanzanalysen für verschiedenartige Zwecke einsetzen, auf unterschiedliche Informationssysteme (i. w. S.) zurückgreifen können und verschiedenartigen situationalen Einflüssen (Zeitdruck etc.) ausgesetzt sind.

b) Empirische Evidenz

Die Frage, ob Studenten "angemessene Stellvertreter" bzw. "surrogates" von Praktikern darstellen, wird zwar in der Literatur kontrovers diskutiert, es liegen aber – im Vergleich zur Anzahl durchgeführter Experimente – nur wenige betriebswirtschaftliche Studien vor, in denen diese Einflüsse empirisch **geprüft** werden.¹⁴⁾

1. Bei unserer Population stellen wir beachtliche **Gemeinsamkeiten** bei den bilanzanalytischen **Urteilen**, auf die sich unser Effizienzmaß "Treffsicherheit der Lagebeurteilung" stützt, mit angehenden Wirtschaftsprüfern und Kreditmanagern fest (vgl. Kapitel 3.4.2.3.). Die Untersuchungen von Knorr

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 337 -
und Weigel belegen erstaunliche Parallelen des **Informationsnachfrageverhaltens** von unseren studentischen Analytikern und Kreditmanagern.¹⁵⁾

Unterschiede erwarten wir bei den Effizienzmaßen "Qualität der Ursachenforschung" und "Frageleistung". Sie setzen erhebliche Erfahrung voraus. Spürbare Lernfortschritte zeigen sich erst nach der dritten und vierten Fallbearbeitung. Daher legen wir unserer empirischen Prüfung nur die Effizienzdimension "Treffsicherheit der Lagebeurteilung" zugrunde. Unsere Konstruktvalidierung belegt, daß die Ursachenanalyse und die Entwicklung qualifizierter Fragen auf einer treffsicheren Lagebeurteilung aufbaut. Wenn das Informationsverhalten keine treffsichere Lagebeurteilung erlaubt, dann ist kaum mit einer anspruchsvollen Ursachenanalyse zu rechnen. Diese Vermutung wird durch pfadanalytische Kontrollauswertungen bestätigt.

2. Die bei unseren Experimenten festgestellten Übereinstimmungen von Studenten der Wirtschaftswissenschaft und Wirtschaftspraktikern sind kein Einzelergebnis. Die in Abbildung 4.8. zusammengefaßten Befunde belegen, daß man nicht zwangsläufig von starken Unterschieden zwischen Studenten und Wirtschaftspraktikern ausgehen muß.¹⁶⁾

Diese Befunde sollen und können nichts "beweisen", denn das Generalisierbarkeitsproblem kann man nicht durch einen induktiven Schluß "lösen". Es gilt auch hier der Satz, daß man **zuerst** begründete Hypothesen aufstellen und dann empirisch überprüfen solle. Wir meinen, es ist Zeit, diese Umkehrung der Beweislast von allen Kritikern zu fordern, die sich **pauschal** gegen Untersuchungen mit Studenten wenden. Nur so kann man die theorielose, simplifizierende Dichotomie "Studenten" vs. "Praktiker" überwinden und zu Studien kommen, die nach den **Ursachen** von Konvergenzen und Divergenzen fragen.

Ein entscheidender Gesichtspunkt scheint uns dabei die Frage zu sein, welche Rolle das schillernde Konstrukt "Erfahrung"¹⁷⁾ spielt. In unseren Experimenten können wir detailliert erforschen, wie sich eine zunehmende Erfahrung auf Informationsverhalten, Effizienz und die zwischen beiden Größen herrschende Beziehung auswirkt. Diese Analysemöglichkeit hat man nur bei Analytikern, die am **Anfang** ihrer Erfahrungskurve stehen:

Abbildung 4.8.: Einflüsse der Dichotomie "Studenten vs. Wirtschaftspraktiker"

Quelle	Aufgabe	Subjekt	Geprüfter Einfluß	Befund
Mock (Information Structures 1969)	Unternehmensspiel, Produktion und Absatz	47 Stud. 25 Manager	Erzielte Gewinne	Keine beachtlichen Unterschiede, laut Angaben im Text S. 146, keine näheren Hinweise, welche Tests durchgeführt werden
Hedberg (Man-Computer-Interaction 1970)	Unternehmensspiel für konkurrierende Banken	62 Stud. 14 Manager	a) Verteilung der Aktivitäten "intelligence", "design", "choice" (S. 169) b) Aktivitätsniveau pro Zeiteinheit (S. 175) c) Entschlußaktivitäten (S. 182) Exp. Bed.: Bildschirm vs. EDV-Liste	a), b) und c): Bei Konstanzhaltung des Angebotsmediums bei beiden Populationen sehr ähnliche Ausprägungen, keine Signifikanztests
Bronner/Witte/Wossidlo (Experimente 1972)	Unternehmensspiel, Produktion, Absatz, Finanzierung,	72 Stud. 72 Manager	a) Anzahl nachgefragter Informationen b) Präzision der Informationsnachfrage Exp. Bed.: "Vitalisierung" der Nachfrage durch 1. Anforderung und 2. Prospekt	a) Haupteffekt n. s., Interaktionseffekte n. s. b) Haupteffekt n. s., Interaktionseffekte n. s.
Hofstedt (Behavioral Parameters 1972)	simulierte Anlageentscheidung	80 Stud. 40 Manager	a) Prognoseleistung Kurs-Gewinn-Verhältnis b) Konfidenzintervall der Prognose c) Zeitverbrauch Exp. Bed.: 1. positive vs. negative wirtschaftl. Entwicklung, 2. konsistente vs. inkonsistente Inform.	a) Haupteffekt n. s., Interaktionseffekte n. s. b) Haupteffekt n. s., Interaktionseffekte n. s. c) Haupteffekt: Praktiker analysieren Bilanzen sign. schneller, Interaktionseffekte n. s.
Bronner (Zeitdruck 1973)	Unternehmensspiel, Produktion, Absatz, Finanzierung,	40 Stud. 72 Manager	Zeitdruck-Sensitivität Exp. Bed.: Variation des Zeitdrucks	Haupteffekt n. s., Interaktionseffekt n. s.
Abdel-khalik (Subject Surrogation 1974)	Kreditvergabeentscheidung	154 Stud. 154 Manager	a) Bewilligter Kreditbetrag für 6 Monate b) "default on loan" binnen 6 Monaten c) "good credit risk" in den nächsten 3 Jahren d) Bewilligter Kreditbetrag für 3 Jahre e) "default on loan" binnen 3 Jahren Exp. Bed.: aggregierte vs. disaggregierte Bilanzdaten	Je 8 U-Tests (4 Firmen, 2 Aggregationsbedingungen, "Table 3"): a) 2 von 8 Tests sign. b) 0 von 8 Tests sign. c) 3 von 8 Tests sign. d) 1 von 8 Tests sign. e) 2 von 8 Tests sign. Die Mittelwerte sind sehr uneinheitlich ausgeprägt ("Table 2"). Daher liegt entgegen der Meinung Abdel-khaliks kein systematischer Unterschied vor.
Moriarity (Graphics 1979)	Beurteilung, ob ein Unternehmen im folgenden Jahr insolvent wird	277 Stud. 20 Manager	Klassifikationsgüte von Unternehmensinsolvenzen Exp. Bed.: "Schematische Gesichter" vs. Kennzahlen	Bei beiden Populationen sign. bessere Leistung mit "Schematischen Gesichtern", kein Interaktionstest
Lucas/Nielsen (Presentation 1980)	Unternehmensspiel, Logistik eines Handelsunternehmens	36 Stud. 78 Manager	Erzielte Gewinne Exp. Bed.: 1. Präsentationsform, 2. Angebotsmedium	bessere Leistungen der Studenten bei erstem Spiel durch sign. höhere Lernfortschritte der Praktiker kompensiert, beim zweiten Spiel gleiche Leistungen, keine Interaktionseffekte getestet
Stock/Watson (Graphics 1984)	Beurteilung, ob ein Bond-Rating im nächsten Jahr steigt, fällt oder gleich bleibt	123 Stud. 14 Manager	Klassifikationsgüte von "Bond-Ratings" Exp. Bed.: "Schematische Gesichter" vs. Kennzahlen	Haupteffekt n. s., Interaktionseffekt n. s.

"The field can be less real than the laboratory because it conceals and abbreviates many basic relationships from which other relationships derive. The field merely reflects the final edited version of quite different processes that happened earlier, more privately, on a smaller scale, and that over time become modified into what we now arbitrarily call the real world. Those earlier processes were and are also real. People who describe the present world have trouble observing those different earlier processes that gave it the shape it now has. The irony, therefore, is that the laboratory now becomes the real world in which one can observe more clearly what is actually going on in concealed form in that more abbreviated artificial world our subjects wander in from."¹⁸⁾

c) Verfügbarkeit von Versuchspersonen

Das Problem "Studenten vs. Praktiker" hat eine nicht zu unterschätzende forschungspragmatische Dimension. Wenn man mehrere komplexe Fälle analysieren lassen will, wird es schwierig, Praktiker zur Mitwirkung zu gewinnen. Die nachfolgende Auswertung experimenteller Studien, in denen Effizienzprüfungen vorgenommen werden, belegt dies (vgl. Abbildung 4.9.). Der Anteil der Untersuchungen, für die Wirtschaftspraktiker als Versuchspersonen gewonnen werden, ist recht niedrig, auch wenn es deutliche Unterschiede zwischen den einzelnen Disziplinen gibt.¹⁹⁾

Ein weiterer Punkt ist bei pragmatischer Würdigung der Experimente zu bedenken: die **Zahl** der Versuchspersonen. Abbildung 4.10. belegt, wieviele Versuchspersonen in empirischen Studien der verhaltensorientierten Rechnungswesenforschung, die das Informationsverhalten **prozessual** erforscht haben, untersucht werden.²⁰⁾

Mit Ausnahme der Studien von Pankoff/Virgil, Weigel und Shields (1984) liegt die Zahl der Versuchspersonen deutlich unter 30. Dies kann darauf zurückgeführt werden, daß sich diese Studien auf die Messung der Informationsbeschaffung beschränken. Bei den geringen Stichprobenumfängen der übrigen Untersuchungen ist es praktisch kaum möglich, statistische Analysen durchzuführen. Selbst bei ausgesprochen starken Effekten in der Grundge-

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 340 -
 samtheit dürfte die Gefahr einer fälschlichen Beibehaltung der Nullhypothese sehr groß sein.²¹⁾ Die meisten Studien gehen diesem Problem aus dem Weg, indem sie – nach gelegentlich heftiger Kritik an der theorielosen bisherigen Forschung – für sich in Anspruch nehmen, lediglich "explorativ" vorzugehen.

Abbildung 4.9.: Art der verwendeten Versuchspersonen in experimentellen Studien mit Effizienzprüfungen

Forschungsfeld	Art der Versuchspersonen			Summe
	nur Studenten	nur Praktiker	Beide Arten	
Computergestützte Informationssysteme	25	2	5	32
Verhaltensorientierte Rechnungswesenforschung	6	5	6	17
Psychologische Forschung zum Lösen komplexer Probleme	16	0	0	16
Summe	47	7	11	65

Die heuristischen Verdienste solcher Studien für die Hypothesenbildung und ein angemesseneres Verständnis realer Probleme sollen nicht geschmälert werden – aber man kann nicht immer auf diesem Stand stehen bleiben. Gerade wenn man das Informationsverhalten in seiner Komplexität würdigen will, braucht man große Stichproben, um komplexe Modelle **prüfen** zu können. Außerdem darf man nicht nur eine Fallbearbeitung vorgeben, wenn man sich – wie in einigen dieser Studien – explizit auf Simon und Newell bezieht, und

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 341 -
hervorhebt, daß die Problemlösung **aufgabenabhängig** sei. Man muß vielmehr **mehrere Fälle** bearbeiten lassen. Dies verschärft das Problem, geeignete Praktiker zu finden.

4.2.4. Design und Ablauf der Experimente²²⁾

Die ausgewerteten Daten werden während der drei Experimentalserien Gießen 82, Kiel 82 und Kiel 84 gewonnen, die alle als zweitägige Blockveranstaltung freitags und samstags abgehalten werden. Am darauffolgenden Montag wird eine Klausur geschrieben, die als Leistungsnachweis bewertet wird. Unsere Auswertungen beziehen sich auf die vier Bilanzanalysen, die unter experimentell kontrollierten Bedingungen angefertigt werden. Der Klausurfall wird nur zur Validierung der Effizienzmessung verwendet. Die Daten der Studenten, die nicht unter experimentellen Versuchsbedingungen arbeiten, werden nicht ausgewertet.

Ca. drei Wochen vor der Blockveranstaltung wird der Wissenstest über Inhalt und Bewertung des Jahresabschlusses durchgeführt. Die Teilnehmer der Serien Gießen 82 und Kiel 82 können außerdem unter Aufsicht einen sog. "Probefall" analysieren. Sie erhalten die aus den Geschäftsberichten fotokopierten Bilanzen und Gewinn- und Verlustrechnungen der Firma Neckermann AG der Jahre 1974 und 1975 und als Vergleichsinformation die Daten der Horten AG von 1974. Der Fall wird als "Versand AG" bezeichnet. Ein Feedback gibt es nicht. Es fehlen die während der Experimente zusätzlich angebotenen Summen- und Saldenpositionen. Von diesem Analyseangebot macht nur ein Teil der Teilnehmer Gebrauch. Bei der Serie Kiel 84 entfällt der Probefall.

Zum Ablauf der Blockveranstaltung schreibt Hauschildt:

"Die Versuchspersonen wurden auf mehrere Räume verteilt und hatten jeweils 90 Minuten Zeit, eine Analyse vorzunehmen. Ihnen wurde dann unmittelbar nach jeder Analyse eine Musterlösung vorgetragen. Dieses Wechselspiel von Analyse und Vortrag der Musterlösung wurde in zwei unmittelbar aufeinanderfolgenden Tagen viermal wiederholt. Die Gruppenzusammensetzungen blieben gleich. Teamarbeit war ausgeschlossen."²³⁾

Abbildung 4.10.: Anzahl der Versuchspersonen in prozessualen empirischen Studien der verhaltensorientierten Rechnungswesenforschung

Studie	Aufgabenstellung	Versuchspersonen
Clarkson (Portfolio Selection 1962)	Anlageentscheidung	1 Portfoliomanager
Pankoff/Virgil (Findings 1970) +	Anlageentscheidung	32 Finanzanalysten
Biggs (Earning Power 1978)	Anlageentscheidung	11 Finanzanalysten
Shields (Information Load 1978)	Analyse von Kosten- und Leistungsberichten	12 Manager
Stephens (Lending Decisions 1980)	Bilanzanalyse für eine Kreditentscheidung	10 Kreditmanager
Weigel (Informationsverhalten 1980) +	Bilanzanalyse für eine Kreditentscheidung	41 Kreditmanager
Stephens et al. (Process Tracing 1981)	Bilanzanalyse für eine Kreditentscheidung	4 Kreditmanager
Anderson (Investment Decision 1982)	Anlageentscheidung	3 Studenten, 4 Finanzanalysten
Bouwman (Expert vs. Novice 1982)	Bilanzanalyse zur Beurteilung der Lage	5 Studenten, 4 Wirtschaftsprüfer
Dillard et al. (Cognitive Analysis 1982)	Buchführungsprobleme	3 Studenten
Biggs/Mock (Auditor Decision 1983)	Prüfung einer Unternehmung	4 Wirtschaftsprüfer
Campbell (Protocol Analysis 1984)	Bilanzanalyse für eine Kreditentscheidung	4 Kreditmanager
Shields (Demand for Information 1984) +	Analyse von Kosten- und Leistungsberichten	36 Studenten
Jacoby et al. (Effectiveness 1985) +	Anlageentscheidung	17 Finanzanalysten

+ Nur die Informationsbeschaffung wird prozeßbegleitend gemessen.

Bei den Experimenten wird die **Präsentationsform** der Daten in folgenden Ausprägungen variiert (vgl. Abbildung 4.11. und 4.12).:

- a) **Präsentation der Daten in Tabellenstruktur:** Die verdichteten Summen- und Saldenpositionen werden in einer tabellarischen (Bilanzstrukturdaten) bzw. staffelartig aufgebauten Form (Erfolgsrechnungsdaten) angeboten.
- b) **Präsentation der Daten in Baumstruktur:** Die verdichteten Summen- und Saldenpositionen werden in einer baumstrukturartigen Form angeboten.
- c) **Präsentation der Daten in Baumstruktur, zusätzlicher Einsatz von baumstrukturen während der zentralen Schulungsveranstaltungen:** Die verdichteten Summen- und Saldenpositionen werden in einer baumstrukturartigen Form angeboten. Zusätzlich werden während der zentralen Schulungsveranstaltungen Baumstrukturen verwendet.²⁴⁾

Zur Variation des experimentellen Faktors schreibt Hauschildt:

"Um die Tabellenstruktur nicht systematisch zu benachteiligen und um den Vergleich realitätsnah zu halten, wurden die Tabellen durch Schriftgrößen, Einrückungen und Umrahmungen einzelner Positionen in sich gegliedert. Die Darstellungsformen - Baumstrukturen versus Tabellen - waren überdies in großen Plakaten in den Arbeitsräumen deutlich sichtbar vor den Versuchspersonen ausgehängt."²⁵⁾

Neben den verdichteten Daten werden die bilanziellen Urdaten in Form des aktienrechtlichen Mindestgliederungsschemas angeboten, einschließlich der Aufgliederung des Anlagevermögens im sog. "Anlagenspiegel". Ausgelassen werden lediglich die sog. "Unter"- und "Vor"-Strichvermerke. Außerdem werden die Positionen II A 1. bis II A 4. des § 151 des Aktiengesetzes zu einer Position "Sachanlagevermögen, Grundstücke und Gebäude" aggregiert. Das reale Informationsangebot wird bewußt **nicht** für experimentelle Zwecke auf wenige Kennzahlen oder eine stark verkürzte Bilanz (i.w.S.) aggregiert, wie in den meisten amerikanischen Untersuchungen, weil damit erhebliche Validitätsprobleme aufgeworfen werden. Daß in unseren Experimenten **zusätz-**

Abbildung 4.11.: Experimentelles Informationsangebot in Baumstruktur

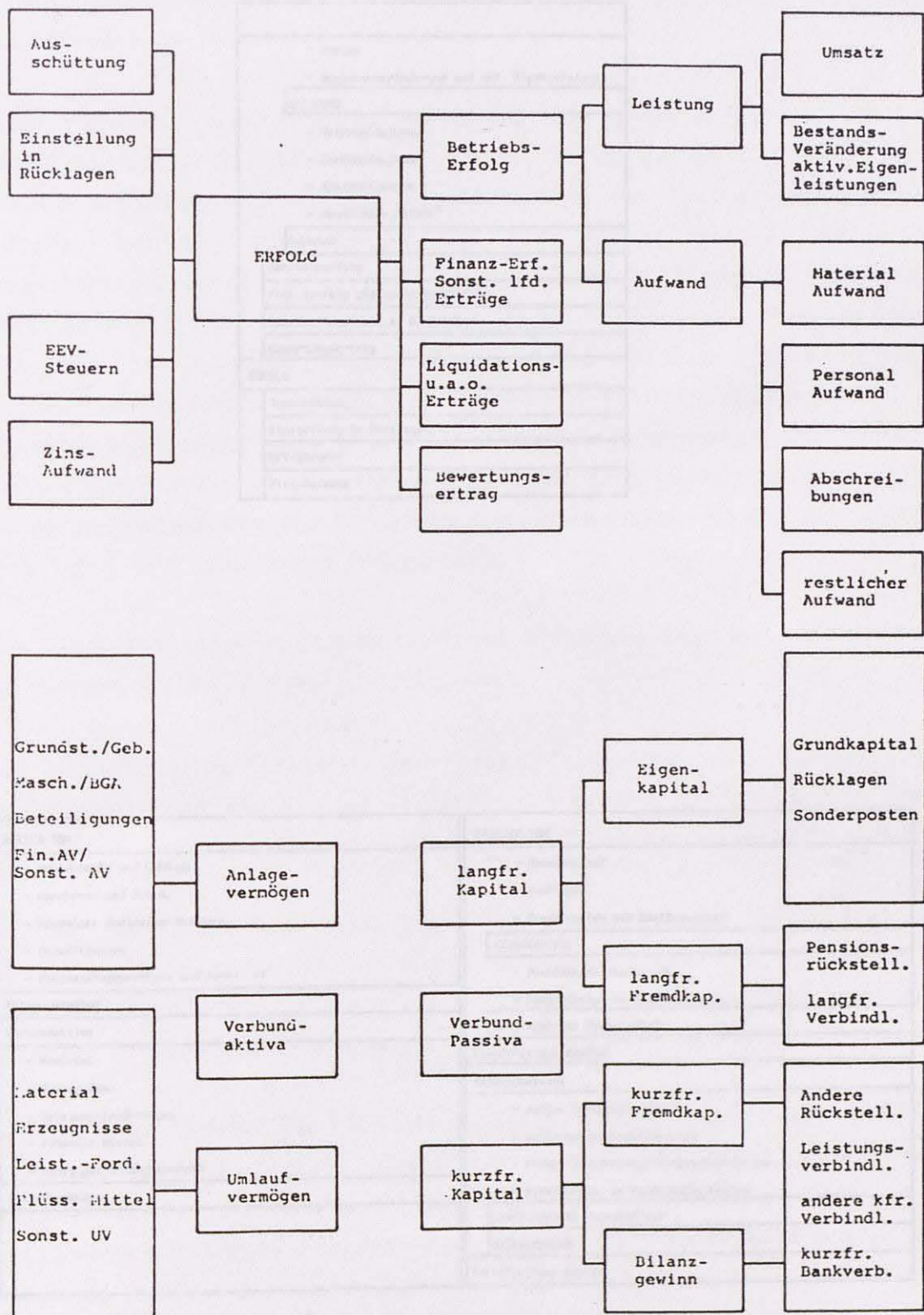


Abbildung 4.12.: Experimentelles Informationsangebot in Tabellenstruktur

- Umsatz
- Bestandsveränderung und akt. Eigenleistung
LEISTUNG
- Materialaufwand
- Personalaufwand
- Abschreibungen
- Restlicher Aufwand
AUFWAND
Betriebserfolg
Finanzerfolg und sonstige Erträge
Liquidations- und a. o. Erträge
Bewertungsertrag
ERFOLG
Ausschüttung
Einstellung in Rücklagen
EEV-Steuern
Zins-Aufwand

AKTIVA TDM	PASSIVA TDM
- Grundstücke und Gebäude	- Grundkapital
- Maschinen und B.G.A.	- Rücklagen
- Sonstiges Sachanlagevermögen	- Sonderposten mit Rücklageanteil
- Beteiligungen	EIGENKAPITAL
- Finanzanlagevermögen und sonst. AV	- Pensionsrückstellungen
Anlagevermögen	- Langfristige Verbindlichkeiten
Verbundaktiva	LANGFRISTIGES FREMDKAPITAL
- Material	Langfristiges Kapital
- Erzeugnisse	Verbundpassiva
- Leistungsforderungen	- Andere Rückstellungen
- Flüssige Mittel	- Leistungsverbindlichkeiten
- Sonstiges Umlaufvermögen	- Andere kurzfristige Verbindlichkeiten
Umlaufvermögen	- Kurzfristige Bankverbindlichkeiten
	KURZFRISTIGES FREMDKAPITAL
	BILANZGEWINN
	Kurzfristiges Kapital

liche Summen- und Saldenpositionen zur Bilanz (i.e.S.) und zur Erfolgsrechnung angeboten werden, ist nach der empirischen Studie von Weigel²⁶⁾ und den Stellungnahmen von Praktikern²⁷⁾ erforderlich, um die externe Validität der Experimente zu sichern.

Die beiden ersten Präsentationsformen werden während der Serien Kiel 82 und Gießen 82 variiert (vgl. Abbildung 4.13.). Die dritte Variante wird nur bei der Experimentalserie Kiel 84 eingesetzt. Diese hat Konsequenzen für die Auswertung, denn die Kandidaten der Serie Kiel 84 besitzen das geringste bilanzanalytische Vorwissen. Die Teilnehmer der Serie Gießen 84 sind Examenkandidaten, die sich auf eine Examensklausur mit einer Bilanzanalyse vorbereiten. Sie haben das höchste bilanzielle Vorwissen. Die Teilnehmer der Serie Kiel 82 haben unmittelbar zuvor bereits an (mindestens) einer Hauptstudiumsveranstaltung über Bilanzierung teilgenommen. Die meisten Teilnehmer der Serie Kiel 84 haben noch keine Hauptstudiumsveranstaltung zum Jahresabschluß besucht.²⁸⁾ Diese Rangfolge deckt sich mit der durchschnittlichen Zahl studierter Fachsemester.

Um die Einflüsse der Präsentationsform und Erfahrung möglichst gut isolieren zu können, bilden wir folgende Kontraste:

- * Zur Bestimmung des Einflusses der **Präsentationsform** werden die der Tabellengruppen der Serien Kiel 82 und Gießen 82 mit den Baumgruppen dieser Serien verglichen. Dieser Kontrast wird **"Baum vs. Tabelle"** genannt.
- * Zur Bestimmung des Einflusses der Erfahrung werden die Serien Kiel 82 und Gießen 82 miteinander verglichen. Dieser Kontrast heißt **"Gießen vs. Kiel 82"**. Anzumerken ist, daß der Einfluß der Erfahrung auch über die intertemporale Auswertung gemessen wird (vgl. Kapitel 4.2.7.).
- * Der Kontrast **"Gießen vs. Kiel 84"** zeigt den **saldierten** Effekt von "Tabelle vs. Baum mit Einsatz während der Schulungsveranstaltung" und "Erfahrung" an. Er kann dazu herangezogen werden, die Einflüsse von Präsentationsform und Erfahrung nochmals zu testen, indem man ihn mit der Summe der beiden übrigen Kontraste vergleicht. Dabei muß man jedoch unterstellen, daß der positive Effekt des zusätzlichen Einsatzes der

Baumstruktur während der Schulung den negativen Effekt der geringeren Erfahrung der Teilnehmer der Serie Kiel 84 gegenüber denen der Serie Kiel 82 gerade kompensiert.

Abbildung 4.13.: Zuordnung der Versuchspersonen zu den Präsentationsformen

Ausprägungen des Faktors Präsentationsform	Experimentelle Serien und Gruppen, Anzahl der Versuchspersonen
tabellarische Darstellung, Schulung ohne optische Strukturierungshilfen	Gießen 82, erste Gruppe (N=17) Kiel 82, erste Gruppe (N=16)
baumstrukturartige Darstel- lung, Schulung ohne opti- sche Strukturierungshilfen	Gießen 82, zweite Gruppe (N=17) Kiel 82, zweite Gruppe (N=17)
baumstrukturartige Darstel- lung, Schulung mit optischer Strukturierungshilfe "Baum"	Kiel 84, erste Gruppe (N=12) Kiel 84, zweite Gruppe (N=12) Kiel 84, dritte Gruppe (N=12)

Innerhalb der Experimentalserien werden die Versuchspersonen den Faktorstu-
fen in folgender Weise zugeordnet:

1. Bei den Serien Gießen 82 und Kiel 82 werden die Versuchspersonen nach dem Alphabet den Gruppen zugeordnet. Dies entspricht einer Zuordnung nach dem Zufallsprinzip und damit den Forderungen, die an Experimente im engeren Sinne gestellt werden.²⁹⁾ Da eine zufällige Zuordnung nur bei hinreichend **großen** Stichproben mit hoher Wahrscheinlichkeit einen zufriedenstellenden Ausgleich personell bedingter, störender Interaktionseffekte garantiert, wird **vor** Durchführung des Experiments geprüft, ob sich die Gruppen bezüglich des Eingangstests zum Bilanzwissen³⁰⁾ signifikant unterscheiden. Dies ist nicht der Fall.

2. Bei der Serie Kiel 84 werden zusätzlich Meßwerte von "Kognitiven Stilen" erhoben.³¹⁾ Da die Versuchsgruppen bezüglich der Kognitiven Stile homogen ausgeprägt sein sollen, werden zunächst Anforderungsprofile erstellt, denen die Gruppen genügen sollten. Dann werden solange Kandidaten aus Grundgesamtheiten mit den verlangten Stilmerkmalen per Zufall zu den Gruppen zugeteilt, bis die Gruppenstärke von 12 erreicht und das Anforderungsprofil erfüllt ist. Signifikante Unterschiede bezüglich des bilanziellen Vorwissens gibt es bei den so gebildeten Gruppen nicht.

4.2.5. Die Validität der Experimente

4.2.5.1. Problemgehalt und Überblick

Das Experiment übernimmt in unserer Untersuchung drei Funktionen:³²⁾

1. Es schafft einen abgeschirmten Rahmen, innerhalb dessen sich das Informationsverhalten unter stets gleichen Bedingungen entfalten kann. Dies gibt uns eine standardisierte, **reproduzierbare Beobachtungsmöglichkeit**, das Informationsverhalten von Analytikern mit unterschiedlichen Eigenschaften zu verschiedenen Zeitpunkten ihrer Erfahrungskurve prozeßbegleitend, differenziert und reliabel messen und zu einer qualitativ gleichwertigen Effizienzmessung in Beziehung setzen zu können.
2. Man kann den gesetzten Rahmen systematisch verändern, indem man gezielt spezifische Parameter des Experimentaldesigns variiert. In dieser Eigenschaft gibt uns das Experiment die Möglichkeit, eine **größere Bandbreite** des Informationsverhaltens abzudecken oder eine bestimmte Bandbreite besonders **intensiv** zu erforschen.
3. Man kann die Versuchspersonen zufällig den variierten Rahmenbedingungen zuordnen. In dieser klassischen Funktion gibt uns das Experiment die Möglichkeit, die **Wirkungen interessierender Gestaltungsparameter kausal zu überprüfen**.

Uns geht es um die Überprüfung der Generalhypothese, daß das Informationsverhalten die Effizienz beeinflußt. Diese Hypothese kann man nicht mit einem Experiment im üblichen Sinne überprüfen, bei dem man einen experimentellen Faktor systematisch variiert und die Versuchspersonen zufällig verschiedenen Ausprägungen dieses Faktors zuordnet.³³⁾ Das Informationsverhalten ist kein beliebig variierbarer experimenteller Faktor, sondern eine intervenierende, nur **indirekt** beeinflussbare Größe, die von **Entscheidungen** des Menschen abhängt. Deshalb sind für unsere Arbeit die ersten beiden Funktionen des Experiments von zentraler Bedeutung. Erklärtes Ziel des von Hauschildt geleiteten Projekts "Informationsgleichgewicht" ist die Einrichtung eines **Informationslabors**. Hierunter sind weniger apparative Ausstattungen als vielmehr ein spezifisches Meßkonzept zu verstehen, mit dem prozessuale Konstrukte angemessen operationalisiert werden.

Wir gehen davon aus, daß das menschliche Informationsverhalten interpersonell, intersituativ und intertemporal so heterogen ausgeprägt ist, daß man nur auf Basis großzahliger, systematisch variierten und replizierter Versuchsreihen zuverlässige Aussagen treffen kann. Daher wurden über einen Zeitraum von fünf Jahren Testversuche, Pretests und Experimente durchgeführt, bevor wir in die eigentliche Auswertungsphase gingen. Im Ergebnis liegen drei vergleichbare Experimentalserien vor, an denen 103 Versuchspersonen, die ein unterschiedliches Vorwissen besitzen, vier unterschiedliche Fälle bearbeiten, unterschiedlichen Präsentationsformen ausgesetzt werden und eine systematische bilanzanalytische Schulung erhalten.

Bevor wir diese Daten auswerten, muß ihre Qualität geprüft werden. Der Prüfprozeß setzte bereits bei der Konzeption und Durchführung der Experimente ein, die **vor** den hier ausgewerteten Serien durchgeführt wurden.³⁴⁾ Es galt in mühevoller Kleinarbeit eine Fülle von Teileinflüssen zu bedenken, empirisch zu prüfen und das Design entsprechend anzupassen. Es würde den Rahmen der vorliegenden Arbeiten sprengen, wollte man alle diese Schritte berichten. Wir beschränken uns auf eine tabellarische Darstellung, die die von Campbell und Mitautoren hervorgehobenen, potentiellen Störfaktoren berücksichtigt³⁵⁾ (vgl. Abbildung 4.14.).

Abbildung 4.14.: Überprüfung der Validitätskriterien für den gewählten Untersuchungsaufbau

Nr	Kriterium und Bedeutung	Relevanz für den Untersuchungsaufbau
1	Zwischenzeitliches Geschehen: Externe Ereignisse, die während der Experimentalseerie auftreten und alternative Erklärungen der beobachteten Effekte ermöglichen	Zeitabstände zwischen den Fallbearbeitungen sehr kurz, keine außerordentlichen Ereignisse aufgetreten, die das experimentelle Verhalten in erkennbarer Weise beeinträchtigt haben
2	Reifung: ein Prozeß innerhalb der Vpn, der Veränderungen nur in Abhängigkeit vom Zeitablauf bewirkt, wie Wachstum, Ermüdung, Hunger, Durst etc.	Beeinträchtigenden Einflüsse einer evtl. Ermüdung sollten sich bei einem Vergleich von vor- und nachmittags bearbeiteten Fällen zeigen. Die gleichmäßige Entwicklung der Meßwerte des Informationsverhaltens und der Effizienz widerlegt diese Vermutung. Die subjektiven Ratings der Vpn (Leistungszufriedenheit, Zeitdruck etc.) zeigen keine signifikanten Unterschiede zwischen den vor- und nachmittags bearbeiteten Fällen.
3	Unzuverlässigkeit von Meßinstrumenten: z. B. (hier nicht relevante Einflüsse): Ermüdung von Beobachtern und Interviewpartnern, Ablesefehler bei technischen Geräten, Störung von Computerlogbüchern durch unterschiedliche Auslastung einer zentral genutzten EDV-Anlage	Einfluß vermutlich sehr gering: a) Instrumente zur Erhebung des Informations- und Beurteilungsverhaltens (für die Effizienzmessung) hoch reliabel und sehr genau; b) Aggregation vieler Informationsaktivitäten eines Prozesses zu Meßwerten der Dimensionen des Informationsverhaltens und der Effizienz bewirkt Ausgleich von Meßfehlern; c) Mittelung der Meßwerte aus vier Analysefällen (bei inter-personellem Vergleich) bewirkt weiteren Ausgleich
4	Testeffekte: Wirkungen, die von der Durchführung eines Tests auf nachfolgendes Verhalten ausgehen	Durchführung eines Eingangstests über bilanzielles Wissen: Kein Einfluß auf die interne Validität, da bei allen Vpn in gleicher Weise durchgeführt. Einfluß auf die externe Validität vermutlich gering, da solche Tests bei Studenten üblich sind und dieser Wissenstest keine Beziehung zur Effizienz der Analyse aufweist. Analyse eines Probefalles: Einfluß auf die interne Validität nicht auszuschließen, da nur bei den Serien Gießen 82 und Kiel 82 angeboten. Einfluß auf exp. Verhalten vermutlich gering, da nur aktienrechtliche Daten angeboten werden und kein Feedback erfolgt.
5	Instrumentierungseffekte: Wirkungen, die von Veränderungen der Meßinstrumente ausgehen	Wird nicht relevant, da Konstanz der Meßinstrumente gesichert. (vgl. Kriterium Nr. 15)
6	Gruppeneffekte: Einflüsse, die auf der Zugehörigkeit zu einer räumlich zusammengefaßten Versuchsgruppe beruhen	Beobachten des Verhaltens anderer an den Information-Displayboards könnte Nachahmer-Effekte auslösen, Versuchsleiter könnten sich unterschiedlich verhalten, Größe, Lage und Ausstattung der Räume könnte – obwohl so weit wie möglich standardisiert – Gruppeneffekte auslösen. Einflüsse statistisch geprüft und als unbeachtlich nachgewiesen (nähere Erläuterung im Text)

Abbildung 4.14.: Überprüfung der Validitätskriterien für den gewählten Untersuchungsaufbau

Nr	Kriterium und Bedeutung	Relevanz für den Untersuchungsaufbau
7	Arbeitsplatzeffekte: Einflüsse, die von unterschiedlichen Arbeitsplätzen herrühren	Ein unterschiedliches Blickfeld könnte die Wirkung der Präsentationsform beeinflussen. Da alle Vpn mit Blickrichtung auf die Information-Display-Boards angeordnet werden, ist dieser Einfluß bereits gedämpft. Unterschiedliche Entfernungen und Gangformen zu den Information-Display-Boards könnte vor allem die Informationsbeschaffung beeinflussen. Störfaktor wird durch Randomisierung ausgeglichen: Arbeitsplätze werden für jede Analyse neu verlost. Statistischer Kontrolltest zeigt keine signifikante Beziehung zwischen Informationsbeschaffung und Arbeitsplatz, so daß sich diese Vorsichtsmaßnahme nachträglich als unbeachtlich erweist.
8	Sog. "Regressions"-Artefakte: Effekte, die durch Auswahl von Personen mit extrem ausgeprägten Eigenschaften entstehen, kann insbes. beim Lernverhalten zu Fehlschlüssen verleiten	Durch Zufallsauswahl aus den Grundgesamtheiten der Serien ausgeschlossen. Bei der Auswertung wird kein Extremgruppenvergleich vorgenommen, auch nicht bei Auswertungen zum Lernverhalten.
9	Auswahlprobleme: Systematische Einflüsse, die von der Anwendung unterschiedlicher Auswahlprinzipien für Vergleichsgruppen herrühren	Wird nicht relevant, da bei allen drei Experimentalserien gleiche Auswahlprinzipien angewandt werden. (Auf das Auswertungsproblem, daß bei der Serie Kiel 84 der Serieneffekt und der Effekt der Präsentationsform konfundiert sind, wird im Text eingegangen).
10	Teilnehmerausfall: Wirkung, die vom unterschiedlichen Ausfall von Teilnehmern auf die Ergebnisse ausgeht.	Entfällt: An den drei Experimenten haben - wie geplant - 103 Versuchspersonen teilgenommen. Sie sollten jeweils vier Fälle während der experimentellen Veranstaltung analysieren. In keiner der 412 geplanten Bearbeitungen gab es einen Ausfall.
11	Zusammenwirken von Auswahl und Reifung (siehe Kriterien Nr. 2 und 9)	Wird nicht relevant, da bei allen drei Serien gleiche Auswahlprinzipien, gleicher Ablauf der Veranstaltung (incl. gleicher Wochentage) und keine verzerrenden Einflüsse durch Ermüdungseffekte.
12	Reihenfolgeeffekte von Tests: Beeinflussung der Ergebnisse eines Tests dadurch, daß die Befragten nach dem Test auf die Variablen sensibler wirken.	Einfluß der Reihenfolge der zu analysierenden Fälle, die in diesem Sinne als "Tests" angesehen werden: Kein Einfluß auf die interne Validität, da bei allen Serien dieselben Fälle in derselben Reihenfolge bearbeitet werden. Externe Validität, Generalisierbarkeit auf das Universum aller möglichen Reihenfolgen: keine Aussage möglich; Randomisierung der Reihenfolge bei den einzelnen Vpn kam wegen Gruppenveranstaltung (Informationsbeschaffung von denselben Information-Display-Boards) und gemeinsamer zentraler Schulungsveranstaltung aller Vpn einer Serie nicht in Frage. Gewählte Reihenfolge entspricht didaktischem Konzept, Fälle mit bisher noch nicht behandelten Problemen bearbeiten zu lassen und von einfachen zu komplexen Fällen vorzugehen. Es stellt sich die Frage, ob eine Generalisierung auf Reihenfolgen, die nicht diesen pädagogischen Grundsätzen entsprechen, überhaupt sinnvoll ist.

Abbildung 4.14.: Überprüfung der Validitätskriterien für den gewählten Untersuchungsaufbau

Nr	Kriterium und Bedeutung	Relevanz für den Untersuchungsaufbau
13	Interaktionseffekte von Auswahl und Test-situation: Die ausgewählten Versuchspersonen reagieren anders auf die Aufgabenstellung als die Personen, auf die man generalisieren will.	Generalisierung auf Wirtschaftspraktiker bleibt ein unsicherer Induktionsschluß, der von uns nicht vorgenommen wird. Wir verweisen auf die ausführliche Diskussion des Problems im Text.
14	Wirkungen des Testaufbaus: Künstlichkeit der Testsituation im Vergleich zu realen Informationssituationen, auf die man generalisieren will	Generalisierung auf (heterogene) Informationsangebots-situationen der Wirtschaftspraxis bleibt ein unsicherer Induktionsschluß. Offen ist auch, inwieweit die Erhebungsmethode "Information-Display-Board" die externe Validität der Experimente beeinträchtigt. Die interne Validität bleibt davon unberührt, weil bei allen Versuchspersonen und Fallbearbeitungen dasselbe Instrument verwendet wird. (Vgl. Diskussion im Text).
15	Sog. "irrelevante Wiederholung" von Experimenten: Problem, ein Experiment in allen seinen Komponenten gleich darzustellen	Durch den Testaufbau gesichert: a) Aufgabenstellung: identische, schriftlich fixierte Aufgabenformulierung, identische Fälle in identischer Reihenfolge; b) Informationsangebot: bis auf experimentell systematisch variierte Präsentationsform identisches Informationsangebot (gleiche Information-Display-Boards, gleiche Zugriffsmodalitäten); c) Versuchsleitereinflüsse (vgl. auch Kriterium 6): striktes Auskunftsverbot über bilanzanalytische Sachfragen, schriftlich fixierte Instruktionen, wie man sich gegenüber Vpn zu verhalten habe, wie die Probeanalyse für das "warm-up" durchzuführen sei, daß sich die Vpn bei der Informationsbeschaffung ungezwungen verhalten sollten, wie die Informationskarten zu signieren seien, der Wechsel der Farbstifte zu erfolgen habe etc.; d) Schulungsveranstaltungen: schriftlich fixierte Vorträge, basierend auf ausführlichen Musterlösungen. Die Vorträge werden bis auf den experimentell variierten Einsatz von optischen Strukturierungshilfen in identischer Form gehalten; e) Arbeitsplätze (vgl. auch Kriterium 7): werden vor jeder Analyse öffentlich verlost, Größe der Arbeitsplätze wird durch Klebestreifen standardisiert; f) Arbeitsmittel: Gleichartig beschaffene Arbeitsmittel (Analysebögen, Stifte, Taschenrechner, etc.) werden in zufälliger Reihenfolge an die Vpn ausgeteilt.

Von besonderer Bedeutung sind Einflüsse, die die Standardisierung der Versuchsbedingungen beeinträchtigen (vgl. hierzu die Kriterien Nr. 1 bis 7, 9 bis 11 und 15). Unter der Rubrik "irrelevante Wiederholung von Experimenten" (Nr. 15) sind Vorkehrungen zusammengefaßt, die getroffen werden, um solche Einflüsse auszuschalten. Die Prüfungen von Gruppen-, Arbeitsplatz- und Ermüdungseffekten, sowie möglicher unzuverlässiger Meßinstrumente werden in den folgenden Abschnitten dargestellt.

4.2.5.2. Die Prüfung von Gruppeneffekten

Für die Prüfung von Gruppeneffekten sprechen erstens theoretische Überlegungen, zweitens empirische Hinweise aus früheren Serien und drittens erhebliche Kostensenkungen gegenüber individuellen Experimenten.³⁶⁾ Viertens müßten die Experimente bei gegebener Kapazität an Versuchsleitern nacheinander durchgeführt werden, was störende Kommunikationsprozesse zwischen den Versuchspersonen auslösen kann.

a) Theoretische Gründe

Die "Sozialpsychologie des Experimentes"³⁷⁾ belegt eine Fülle möglicher Störvariablen. Außerdem können physische Umstände, wie z. B. Einflüsse des Versuchsraums eine Rolle spielen. Nach der Untersuchung von Silberer und Frey kann darüber hinaus mit einem designspezifischen "Nachahmer"-Effekt bei der Informationsbeschaffung gerechnet werden, wenn eine Bezugsperson beobachtet wird, die besonders viele Informationen nachfragt.³⁸⁾

b) Hinweise anderer Untersuchungen

Silberer und Frey prüfen diese These in einem Individualexperiment, indem sie eine zweite Versuchsperson als Verhaltensmodell "in einer für die Vp erkennbaren Weise"³⁹⁾ eine besonders große Menge von Informationen beschaffen lassen (14 von 15 möglichen). Außerdem wird der Zeitpunkt des "modelling" variiert: Die Bezugsperson kann vor oder nach Kenntnisnahme des Informationsangebots beobachtet werden. Die Informationsbeschaffung steigt von 9,13 auf 10,47 Informationen signifikant an, wobei der Wert von 10,47

bei beiden Versuchsbedingungen mit Beobachtungsmöglichkeit erzielt wird. Trotz des hohen Aufforderungscharakters der dyadischen Versuchssituation ist der Effekt schwach ausgeprägt: Die erklärte Varianz beträgt 4,9 v.H.⁴⁰⁾

Knorr untersucht die Informationsbeschaffungsmengen der Serien Gießen 82 und Kiel 82. Er stellt fest, daß beide Populationen auf die gleichen Präsentationsformen unterschiedlich reagieren. Er führt dies auf gruppendynamische Effekte zurück.⁴¹⁾ Da sich die Einflüsse der Gruppenzugehörigkeit nicht von den Haupt- und Interaktionseffekten von Hochschulort und Präsentationsform isolieren lassen, ist seine Schlußfolgerung nicht zwingend. Gegen Knorrs Vermutung spricht die Tatsache, daß sich die Standardabweichungen innerhalb der Gruppen unregelmäßig entwickeln.⁴²⁾ Bei einer Orientierung am Verhalten anderer Gruppenmitglieder sollten sie abnehmen.

c) Empirische Prüfung

Wenn es Gruppeneffekte gibt, dann sollten sie sich besonders deutlich bei den 144 Informationsprozessen der Serie Kiel 84 zeigen, weil dort **kein** experimenteller Faktor variiert wird und alle Versuchspersonen aus **derselben** Population stammen. Wir prüfen dies durch einen varianzanalytischen Test, der sich nicht nur auf die Informationsbeschaffung, sondern auf alle sieben Verhaltensdimensionen unseres theoretischen Modells bezieht.

Der Befund ist eindeutig: Bei keiner Verhaltensdimension werden gängige Signifikanzniveaus erreicht. Die Schätzung für die erklärte Varianz in der Grundgesamtheit (Hays' Omega) ist meistens negativ⁴³⁾ (vgl. Tabelle 4.2.). Bei der Informationsbeschaffung beträgt sie 0,62 v. H. Dieser geringe Wert steht in keinem Verhältnis zu dem hohen Nutzen von Gruppenexperimenten.

4.2.5.3. Die Prüfung von Arbeitsplatzeffekten

Arbeitsplatzeffekte werden dadurch reduziert, daß die Arbeitsplätze vor jeder Analyse öffentlich verlost werden (weitere Angaben in der tabellari-schen Übersicht). Wenn man die Meßwerte einer Versuchsperson über mehrere Bilanzanalysen mittelt – und dies tun wir – dann dürften sich diese Ein-flüsse ausgleichen. Trotzdem kann es für Auswertungen von einzelnen Analy-

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 355 -
 sefallen interessant sein, wie groß die Arbeitsplatzeffekte sind. Unsere
 Prüfung zeigt keinen signifikanten Einfluß.⁴⁴⁾

Tabelle: 4.2.: Ergebnisse der Prüfung von Gruppeneffekten

Verhaltensdimension	Varianz zwischen den Gruppen		Signi- fikanz- niveau
	Stichprobe (Eta)	Grundgesamt- heit (Omega)	
Verrichtungsgliederung	0,47 %	negativ	0,72
Objektgliederung	0,41 %	negativ	0,75
Vergleichstechnik	1,59 %	0,15 %	0,32
Aggregationsgrad	0,67 %	negativ	0,62
Beschaffung	2,61 %	0,62 %	0,15
Übertragung	0,37 %	negativ	0,97
Kennzahlenzahlenbildung	0,16 %	negativ	0,90

4.2.5.4. Die Prüfung von Ermüdungseffekten

Die Experimentalserien werden als Blockveranstaltung durchgeführt. Es finden vor- und nachmittags je eine 90-minütige Bilanzanalyse und eine gleich lange zentrale Schulungsveranstaltung statt. Führt diese kompakte Art der Veranstaltung zu Ermüdungseffekten? Kommt es während der Nachmittagsveranstaltungen zu einem Leistungsabfall?

1. Betrachtet man die Entwicklung der Niveaus der Informationsaktivitäten und die der Effizienzmeßwerte, so kann man einen starken Anstieg erkennen, der einer recht gleichmäßig, degressiv steigenden Lernkurve entspricht und damit nennenswerte Ermüdungseffekte ausschließt.⁴⁵⁾

2. Außerdem wird geprüft, ob sich die subjektiv **wahrgenommene** Beurteilungssituation zwischen Vor- und Nachmittagsitzungen signifikant unterscheidet. Wir können uns dabei auf die nachfolgend wiedergegebenen Likertskalen, die nach jeder Bilanzanalyse erfragt werden, stützen. Obwohl sich

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 356 -
dieser Test auf wenigstens 410⁴⁶⁾ Informationsprozesse stützt ergeben sich bei keinem Rating signifikante Unterschiede (vgl. Tabelle 4.3.). Rechnet man den t-Wert der Zeitdruck-Skala, die das geringste Signifikanzniveau aufweist, in einen punktbiserialen Korrelationskoeffizienten um, ergibt sich ein Wert von 0,07.⁴⁷⁾ Nennenswerte Beeinflussungen des Ergebnisses durch Ermüdungseffekte können daher ausgeschlossen werden.

Tabelle 4.3.: Ergebnisse der Prüfung wahrgenommener Ermüdungseffekte
(mittlere Skalenwerte, Standardabweichungen in Klammern darunter)

Zu beurteilendes Likertskalen-Item	vor- mittags	nach- mittags	Signifi- kanzniv.
"Ich bin <u>sicher</u> , die Unternehmung richtig beurteilt zu haben."	4,6 (1,6)	4,7 (1,6)	0,49
"Ich bin der Meinung, daß ich genügend Informationen <u>nachgefragt</u> habe."	4,1 (1,6)	4,2 (1,6)	0,42
"Zur Lösung der Aufgabe benötige ich keine <u>weitergehenden Informationen</u> ."	4,5 (1,9)	4,7 (1,8)	0,31
"Ich bin mit dem <u>Verlauf</u> meiner letzten Analyse sehr zufrieden."	4,9 (1,5)	4,8 (1,6)	0,92
"Ich bin mit meinem <u>Analyseergebnis</u> sehr zufrieden."	5,0 (1,5)	5,0 (1,5)	0,90
"Ich hatte für die Analyse genügend <u>Zeit</u> zur Verfügung."	5,3 (1,9)	5,5 (1,7)	0,14

Pole der Likert-Skalen: (1 "stimme völlig", 7 "stimme überhaupt nicht zu")

4.2.6. Die Messung des Informationsverhaltens

4.2.6.1. Die Erhebungsmethoden

In den Experimenten werden folgende Erhebungsmethoden angewandt:

1. Erhebung des Informationsbeschaffungsverhaltens mit Hilfe von Information-Display-Boards.
2. Erhebung der Beurteilungsaktivitäten mit Hilfe einer eigens entwickelten Variante der Inhaltsanalyse.
3. Erhebung der "übrigen" Informationsaktivitäten mit Hilfe einer weiteren, eigens entwickelten Variante der Inhaltsanalyse.
4. Erhebung von subjektiven Beurteilungen mit Hilfe von Fragebögen, die nach jeder Analyse an die Versuchspersonen verteilt wurden.
5. Erhebung des bilanziellen Wissens mit Hilfe eines Eingangstests.

Im vorliegenden Gliederungsabschnitt wird die Erhebung der sog. "übrigen" Informationsaktivitäten zusammenfassend dargestellt und ihre Reliabilität überprüft. Eine ausführliche Darstellung findet sich in dem Forschungsbericht von Gemünden und Petersen.⁴⁸⁾ Die Messung der Beschaffungsaktivitäten ist Gegenstand der Dissertation von Knorr.⁴⁹⁾ Auf die Beurteilungsaktivitäten wurde in Kapitel 3.2.4. eingegangen.

4.2.6.2. Die operationalen Definitionen der Informationsaktivitäten

Für die Informationsaktivitäten gelten folgende operationalen Definitionen:

1. Unter einer **Informationsbeschaffungsaktivität** verstehen wir das Ziehen eines Informationskärtchens aus einer Tasche eines Information-Display-Boards.

2. Als **"Übertragungs- und Speicherungsaktivitäten"** werden folgende Vorgänge angesehen:

- (a) das Übertragen einer beschafften numerischen Jahresabschlußinformation von einem Informationskärtchen auf den Analysebogen,
- (b) das Übertragen einer vom Analytiker logisch erschlossenen numerischen Jahresabschlußinformation auf den Analysebogen.⁵⁰⁾
- (c) das Übertragen einer bereits auf dem Analysebogen gespeicherten numerischen Jahresabschlußinformation an eine andere Stelle des Analysebogens,
- (d) das Übertragen einer bereits auf dem Analysebogen gespeicherten, vom Analytiker errechneten Kennzahl, an eine andere Stelle des Analysebogens,

Bei jedem dieser vier Vorgänge ist pro übertragener Zahl eine Übertragungs- und Speicherungsaktivität zu codieren.

3. Als **Rechenaktivitäten** (synonym: **Kennzahlenbildungen**) werden alle arithmetischen Operationen verstanden, mit denen aus wenigstens zwei beschafften numerischen Jahresabschlußinformationen oder bereits aus ihnen errechneten Kennzahlen eine dritte numerische Information abgeleitet wird. Dabei ist pro errechneter und auf dem Analysebogen gespeicherter Zahl, unabhängig davon, ob es sich um ein Zwischen- oder Endergebnis handelt, eine Rechenaktivität zu codieren.

Die Anzahl der bei dieser Aktivität angewandten arithmetischen Operationen ist die **Intensität** der Rechenaktivität. Sie wird als Merkmal einer Rechenaktivität codiert.

4.2.6.3. Die Prüfung der Reliabilität der Informationsaktivitäten

Zur Bestimmung der Reliabilität der Informationsaktivitäten werden die in Kapitel 3.2.4. dargestellten Maßzahlen verwendet. Für unsere Prüfung ergeben sich hohe Werte, die deutlich über denen vergleichbarer Untersuchungen liegen (vgl. Tabelle 4.4. und Tabelle 3.12.). Dies kann vor allem darauf zurückgeführt werden, daß der **Ermessensspielraum** bei der Codierung außerordentlich **gering** ist. In der großen Mehrheit der Fälle läßt sich zweifelsfrei feststellen, welche Informationen übertragen bzw. welche Kennzahlen gebildet werden.

Tabelle 4.4.: Reliabilitätskennzahlen der Informationsaktivitäten

Informationsaktivität	Identifikationsrate	Übereinstimmungsrate	Cohens Kappa
Übertragungs- und Speichereungsaktivität	0,979	0,989	0,99
Rechenaktivität (synonym: Kennzahlenbildung)	0,976	0,976	0,98

4.2.6.4. Die operationalen Definitionen der Dimensionen des Informationsverhaltens

Die solchermaßen erhobenen und geprüften Informationsaktivitäten werden nun zu Skalen verdichtet. Die Berechnungsvorschriften sind auf der nächsten Seite zusammengefaßt.

$$\text{Index der Objektgliederung "Erfolgslage" vor "Vermögens- und Finanzlage"} = \frac{E \text{ vor VF} - VF \text{ vor E}}{E \text{ vor VF} + VF \text{ vor E}}$$

dabei bedeuten:

E vor VF: Anzahl der Reihenfolgebeziehungen, in denen eine Informationsaktivität zum Objekt "Erfolgslage" vor einer Informationsaktivität zum Objekt "Vermögens- und Finanzlage" stattfindet.

VF vor E: Anzahl der Reihenfolgebeziehungen, in denen eine Informationsaktivität zum Objekt "Vermögens- und Finanzlage" vor einer Informationsaktivität stattfindet zum Objekt "Erfolgslage" stattfindet.

$$\text{Index der Verrichtungsgliederung} = \frac{A \text{ vor B} - B \text{ vor A}}{A \text{ vor B} + B \text{ vor A}}$$

dabei bedeuten:

A vor B: Anzahl der phasentheoremkonformen Reihenfolgebeziehungen der Verrichtungen

B vor A: Anzahl der phasentheoremkonträren Reihenfolgebeziehungen der Verrichtungen

Phasentheoremkonforme Folge der Verrichtungen: Beschaffung, Übertragung und Speicherung, Primäre Bewertung, Kennzahlenbildung, Beurteilung, Fragen

$$\begin{aligned} & \frac{\text{Anzahl beschaffter Daten zum Vergleichsunternehmen}}{\text{Anzahl aller beschafften Daten}} * 50,0 \\ + & \frac{\text{Anzahl übertragener Daten zum Vergleichsunternehmen}}{\text{Anzahl aller übertragenen Daten}} * 50,0 \\ = & \text{Komplexität der Vergleichstechnik} \end{aligned}$$

$$\begin{aligned} & \frac{\text{Anzahl beschaffter verdichteter Daten}}{\text{Anzahl aller beschafften Daten}} * 33,3 \\ + & \frac{\text{Anzahl übertragener verdichteter Daten}}{\text{Anzahl aller übertragenen Daten}} * 33,3 \\ + & \frac{\text{Anzahl Kennzahlen, die aus verdichteten Daten gebildet werden}}{\text{Anzahl aller gebildeten Kennzahlen}} * 33,3 \\ = & \text{Aggregationsgrad der Informationsaktivitäten} \end{aligned}$$

4.2.6.5. Die Prüfung der Validität der Dimensionen des Informationsverhaltens

In den vorangegangenen Gliederungsabschnitten haben wir zwei der vier Validitätsarten von Cook und Campbell behandelt: die interne und die externe Validität. Nun ist die "construct validity" zu prüfen. Für die **Effizienzmessung** wurde sie bereits in Kapitel 3.4. nachgewiesen. Wir beschränken uns daher auf das Informationsverhalten.

Konstruktvalidität fordert die gleichzeitige Erfüllung von Konvergenz-, Diskriminanz- und Nomologischer Validität.⁵¹⁾ Diese Anforderungen werden folgendermaßen geprüft:

1. **Nomologische Validität** verlangt die Prüfung theoretisch abgeleiteter Beziehungen zu einem Außenkriterium. Dieses Kriterium ist für uns die Effizienz der Bilanzanalyse. Die Prüfung der Nomologischen Validität fällt daher mit dem Test unserer Effizienzhypothesen zum Informationsverhalten zusammen (vgl. Kapitel 4.3.).
2. **Diskriminanzvalidität** besagt, daß Indikatoren, die verschiedenen Konstrukten zugeordnet werden, niedriger miteinander korrelieren als mit solchen, die demselben Konstrukt zugeordnet werden. Diese Anforderung wird durch die Prüfung der Hypothesen zum Zusammenhang der verschiedenen Dimensionen des Informationsverhaltens getestet (vgl. Kapitel 4.3.). Wir hatten postuliert, daß die Dimensionen, die verschiedenen Ebenen zugeordnet werden, voneinander unabhängige Gestaltungsparameter darstellen. Bei Dimensionen der gleichen Verhaltensebene hatten wir bestimmte Abhängigkeiten abgeleitet.
3. **Konvergenzvalidität** besagt, daß Indikatoren, die demselben Konstrukt zugeordnet werden, signifikant und hoch miteinander korrelieren. Diese Anforderung wird im vorliegenden Kapitel geprüft.

Zuvor wird die **Inhaltsvalidität** der Messung erörtert.

a) Inhaltsvalidität

Unter Inhaltsvalidität verstehen wir, daß das **tatsächliche** Informationsverhalten hinreichend differenziert gemessen wird. Die gebildeten Skalen sollen die intendierten Aspekte erfassen. Das Informationsverhalten soll durch die Messung nicht beeinflusst werden. Außerdem gehört hierzu die bereits geprüfte Anforderung, daß sich das gemessene Informationsverhalten auf ein realistisches Informationsangebot bezieht.

1. Prozeßbegleitende Messung: In unseren Experimenten wird das tatsächliche Informationsverhalten prozeßbegleitend erhoben. Sie unterscheiden sich damit von der großen Mehrheit der Studien zum Informationsverhalten, die sich auf ein **nachträglich erfragtes** Verhalten bezieht. Eine Auswertung von 100 Forschungsergebnissen zum Zusammenhang von Informationssuche und wahrgenommenem Risiko ergibt folgende Verteilung⁵²⁾ (vgl. Abbildung 4.15.):

Abbildung 4.15.: Messung des Informationsverhaltens in Marketingstudien

Art der Messung des Informationsverhaltens	Anzahl der Befunde
Retrospektiv erfragtes Verhalten: (Fragebogen, Interview)	60
Prospektiv erfragtes Verhalten: (Fragen, bezogen auf hypothetische Situationen)	10
Prozeßbegleitend erhobenes Verhalten: (vornehmlich Information-Display-Board)	30

Dieses Ergebnis spiegelt den Stand der **Marketingforschung** wider. Anzumerken ist, daß die wenigen Studien, die nicht nur die Informationsbeschaffung, sondern auch die -verarbeitung untersuchen, nicht Gegenstand der hier referierten Meta-Analyse sind.⁵³⁾

In der **verhaltensorientierten Rechnungswesenforschung** sieht das Bild noch ungünstiger aus. Nach den "State-of-the-Art"-Artikeln von Libby und Lewis ergibt sich die in Abbildung 4.16. dargestellte Verteilung.⁵⁴⁾ **Der Befund belegt, daß nur ein Bruchteil der Studien das Informationsverhalten prozeßbegleitend erforscht.**⁵⁵⁾

Abbildung 4.16.: Messung des Informationsverhaltens in der Rechnungswesenforschung

Art der Messung des Informationsverhaltens	Empirische Studien	
	abs.	v. H.
Durch faktorielles Design und statistische Analyse ermittelte Informationsgewichtung (sog. "lens-model studies")	37	43,0
Schätzung subjektiver Wahrscheinlichkeiten (sog. "probabilistic judgment studies")	29	33,7
Kognitive Stile (werden zu versch. Verhaltensdimensionen und Effizienz in Beziehung gesetzt)	14	16,3
Prozeßbegleitend erhobenes Verhalten (sog. "predecisional behavior-studies")	6	7,0

2. Vollständige und differenzierte Messung: Unser Konzept erfaßt nicht nur die Informationsbeschaffung, sondern auch die Übertragung, Kennzahlenbildung, Beurteilung und Fragen. Die Messung ist **vergleichsweise** vollständiger als die der meisten anderen Untersuchungen. Die Feinheit der Objektmessung und die Reliabilität der Reihenfolgemessung wird in den Dokumentationen der Erhebungsinstrumente belegt.⁵⁶⁾ Um das zeitliche Auftreten der Aktivitäten messen zu können, werden alle 15 Minuten neue Faserstifte in einer anderen Farbe ausgeteilt. Daher ist es möglich, den 90-minütigen Informationsprozeß nach sechs Zeittakten zu differenzieren.⁵⁷⁾

3. Keine Beeinflussung des Informationsverhaltens durch das Meßverfahren:

Die Bewertung dieser Anforderung ist nach den beiden verwendeten Verfahren "Inhaltsanalyse der schriftlichen Ausarbeitungen" und "Information-Display-Board-Messung der Informationsbeschaffung" zu differenzieren.

Die **Inhaltsanalyse**⁵⁸⁾ kann als nicht-reaktives Instrument angesehen werden, insbesondere im Vergleich zu anderen Erhebungsmethoden: Gegenüber dem Denke-Laut-Protokoll entfällt das Problem, eine Zusatzaufgabe anzufertigen. Beeinträchtigende Wirkungen durch eine Selbstreflexion des Verhaltens werden vermieden. Störeffekte durch die soziale Interaktion von Versuchsperson und Versuchsleiter entfallen. Es gibt keine Interaktionsbarrieren der Mensch-Maschine-Kommunikation wie beim Computer-Logbuch-Verfahren, keine Beeinträchtigungen durch ungenaue Apparate wie bei der Blickaufzeichnungskamera. Es fehlt der bei diesem Instrument nicht von der Hand zu weisende Einfluß der Laborsituation, daß man genau beobachtet wird.

Das **Information-Display-Board** ist ungünstiger einzuschätzen: Der motorische Zugriff auf viele einzelne Informationskärtchen und das Erfordernis, diese subjekt- und problemgerecht zu ordnen, stellen andere Anforderungen an das Informationsverhalten als ein Wandern des Blicks über ein Formular.⁵⁹⁾ Aus pädagogischer Sicht ist jedoch positiv zu vermerken, daß die Informationsbeschaffung von den Boards das Erlebnis der Informationsüberlastung bewußt macht und dazu anregt, Informationen gezielter zu beschaffen.⁶⁰⁾

4. Angemessene Skalenbildung

"Inhaltsvalidität fordert ... die präzise semantische Abgrenzung eines Konstrukts und eine Festlegung aller seiner Facetten."⁶¹⁾ Dies bedeutet, daß man die Konstrukte nicht "zu eng" definieren darf.⁶²⁾ Diese Anforderung ist nach den Meßwerten der einzelnen Dimensionen des Informationsverhaltens zu differenzieren.

Den Skalen zur Gliederung des Prozesses nach Objekten und Verrichtungen werden alle inhaltsanalytisch⁶³⁾ gemessenen Arten von Informationsaktivitäten zugrundegelegt, nicht nur die Informationsbeschaffung, wie bei den meisten empirischen Studien, die Reihenfolgemaße bilden.⁶⁴⁾ Die Transitionsmaße dieser Studien berücksichtigen nur Reihenfolgebeziehungen zu direkten Nachfolgern. Petersens Meßwerte berücksichtigen die Sequenzbeziehungen zu **allen** Nachfolgern.⁶⁵⁾

Die Skalen zum Aggregationsgrad bzw. zur Vergleichstechnik stützen sich nicht nur auf die Informationsbeschaffung, sondern auch auf die Übertragung und Speicherung sowie die Kennzahlenbildung.⁶⁶⁾

Bei den operativen Skalen kann man davon ausgehen, daß die Mengen der beschafften und übertragenen Daten sowie der berechneten Kennzahlen nahezu vollständig gemessen werden, so daß auch bei diesen Dimensionen die Operationalisierung dem Meßziel entspricht.⁶⁷⁾

b) Konvergenzvalidität

Konvergenzvalidität bezeichnet nach Hildebrandt den "Grad, zu dem zwei maximal unterschiedliche Messungen des gleichen Konstrukts in ihren Werten übereinstimmen."⁶⁸⁾ Diese Definition ist nicht ganz unproblematisch, weil es schwer sein dürfte, **a priori** zu entscheiden, wann zwei Messungen gerade noch das gleiche Konstrukt messen, so daß sie **maximal** unterschiedlich sind. Wir prüfen daher, ob Maßgrößen, die mit unterschiedlichen Erhebungsverfahren gewonnen werden, aber dasselbe Konstrukt messen sollen, zufriedenstellend konvergieren. Außerdem prüfen wir, ob es konsistente interindividuelle Verhaltensunterschiede über verschiedene Fälle hinweg gibt. (Für die Effizienzmessung wird diese Eigenschaft in Kapitel 3.4.4. nachgewiesen).

1. Zur Prüfung der Konvergenzvalidität der Skala **Verrichtungsgliederung der Informationsaktivitäten** wird diese mit einem anders gebildeten Meßwert, dem sogenannten **Prozeßspektrenwert** von Petersen, verglichen. Er bezieht auch die Informationsbeschaffungsaktivitäten mit ein, berücksichtigt dafür aber lediglich die Ordnung der Informationsaktivitäten nach den 6 Zeittakten des Prozesses, die durch die farbliche Kennzeichnung der signierten Informa-

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 366 -
tionskarten und der beschriebenen Analysebögen identifiziert werden können.⁶⁹⁾ Die Korrelation zwischen beiden Meßwerten beträgt 0,82 und bestätigt damit eindeutig die Konvergenzvalidität dieser Skala.⁷⁰⁾

2. Zur **Reihenfolge der Bearbeitung der Objekte** wurde (noch) kein "Prozeßspektrenwert" von Petersen vorgelegt. Wir können aber prüfen, ob ein Analytiker, der die Analyse mit der Erfolgslage beginnt, auch einen höheren Anteil von Informationen zur Erfolgslage beschafft, überträgt und verarbeitet. Die entsprechenden Korrelationen betragen 0,42, 0,50 und 0,35. Sie bestätigen damit die Konvergenzvalidität. Mit einem gleich hohen Wert wie bei der ersten Prüfung durften wir nicht rechnen, denn ein Meßwert zum Prozeßverlauf erfaßt nicht die gleichen Aspekte wie ein Meßwert zur inhaltlichen Struktur. Vergleichbare Konvergenzkoeffizienten aus der Marketingforschung liegen "zwischen 0,20 und 0,60, wobei studentische Populationen engere Korrelationen aufweisen als nichtstudentische."⁷¹⁾ Für die verhaltensorientierte Rechnungswesenforschung gilt dieser Populationseffekt nicht. Für Manager, die sog. "performance reports" beurteilen, bestimmt Shields eine mittlere Rangkorrelation von 0,46.⁷²⁾ In einer Nachfolgestudie mit Studenten liegen die Meßwerte zwischen 0,12 und 0,22.⁷³⁾

3. In die Skala **Komplexität der Vergleichstechnik** gehen die Subskalen "Komplexität der Vergleichstechnik bei der Informationsbeschaffung" und "Komplexität der Vergleichstechnik bei der Übertragung und Speicherung" ein. Beide korrelieren mit 0,66 und bestätigen damit die Konvergenzvalidität.

4. Die Skala **Aggregationsgrad der Informationsaktivitäten** stützt sich auf die Komponenten Beschaffung, Übertragung und Speicherung sowie Kennzahlenbildung. Cronbachs Alpha für die drei Subskalen beträgt 0,60.⁷⁴⁾

5. Zur Bestimmung der Konsistenz interindividueller Unterschiede wird ebenfalls Cronbachs Alpha verwendet. Wenn zwei Stichproben von Testaufgaben zufällig aus einem Pool von Testaufgaben gezogen werden, dann gibt Cronbachs Alpha die erwartete Korrelation zwischen den summierten Meßwerten beider Testserien an.⁷⁵⁾ In unserem Falle ist Alpha als deskriptives Maß anzusehen, da die vier Testaufgaben nicht systematisch ausgewählt werden und auch ihre Reihenfolge nicht zufällig ist. Die in Tabelle 4.5.

dokumentierten Meßwerte sind als hoch anzusehen, wenn man bedenkt, daß hier nicht zahlreiche "Items" zugrundegelegt werden, die sich nur in semantischen Nuancen unterscheiden, sondern komplexe Analysefälle. **Man kann die Verhaltensdimensionen über mehrere Fälle hinweg generalisieren.**

Tabelle 4.5.: Prognosevalidität der Dimensionen des Informationsquenzen für die Anwand verhaltens

Verhaltensdimension	Cronbachs Alpha
Verrichtungsorientierung	0,66
Objektorientierung	0,59
Vergleichstechnik	0,89
Aggregationsgrad	0,72
Informationsbeschaffung	0,83
Übertragung	0,89
Kennzahlenbildung	0,79

4.2.7. Die statistischen Auswertungsverfahren

4.2.7.1. Das Auswertungsprogramm

Unserer Untersuchung liegt eine probative Fragestellung zugrunde: Ein gegebenes Hypothesensystem soll getestet werden. Explorative, heuristische Verfahren der Datenanalyse wie z. B. Clusteranalysen oder **schrittweise** Diskriminanzanalysen sind daher ausgeschlossen. Es werden vielmehr Pfadanalysen mit Hilfe multipler Regressionsanalysen durchgeführt. Dies entspricht der Art der Hypothesenformulierung. Dies bedeutet nicht, daß nur ein Test durchgeführt wird. Man kann die bewußt allgemein formulierten Hypothesen unterschiedlich interpretieren und spezifische Tests durchführen, die jeweiligen Hypothesenversionen strenger prüfen.

a) Differenzierung nach der Art der Fragestellung

Nach der Art der Fragestellung ist zu unterscheiden, ob interpersonelle oder intertemporale Zusammenhänge analysiert werden, oder ob man eine kombinierte Analyse durchführt, die auch Wechselwirkungen berücksichtigt. Abbildung 4.17. erläutert diese Fallunterscheidungen. Sie zeigt deren Forschungsfragen, Hypotheseninterpretationen, Meßwertbildungen und ihre Konsequenzen für die Anwendung.

1. Bei der **interpersonellen Analyse** werden die Hypothesen als interindividuelle Unterschieds- oder Zusammenhangsaussagen interpretiert. Um diese Hypothesen zu testen, werden die Meßwerte der Analytiker aus den vier Fällen gemittelt und zu den durchschnittlichen Effizienzmeßwerten der vier Fälle in Beziehung gesetzt. Eine derartige Regressionsanalyse mit Mittelwerten ist sinnvoll, weil die sie das typische Verhalten eines Analytikers zuverlässiger beschreiben als eine einzelne Messung.

2. Bei der **intertemporalen Analyse** wird das **Lernverhalten** untersucht. Diese Prüfung geht von einer dynamischen Interpretation der Hypothesen aus. Um sie zu testen, werden **ipsative Meßwerte** gebildet.⁷⁶⁾ Man erhält sie, indem man von den individuellen Mittelwerten der Analytiker, gebildet über die vier Fälle, die Meßwerte der einzelnen Fälle abzieht. Dadurch werden die unterschiedlichen Niveaus der einzelnen Versuchspersonen, die bereits in der interpersonellen Analyse betrachtet wurden, eliminiert.

3. Bei einer **kombinierten Zeitreihen- und Querschnittsanalyse** können mehrere Schätzmethoden angewandt werden. Da im Vergleich zu ökonometrischen Zeitreihen nur sehr wenige Meßwiederholungen vorliegen, beschränken wir uns auf ein einfaches, leicht nachvollziehbares Verfahren. Wir testen unser Hypothesensystem für jeden Analysefall und erhalten eine Bandbreite von Regressionskoeffizienten, innerhalb derer die wahren Parameter liegen. In einem weiteren Schritt wird geprüft, ob es Interaktionswirkungen von Informationsverhalten und Falleinfluß gibt.

Abbildung 4.17.: Die Art der untersuchten Fragestellungen

	Interpersonelle Unterschiede	Intertemporale Unterschiede	Kombinierte Betrachtung
Forschungsfrage	Besitzen gute Analytiker ein anderes Informationsverhalten als schlechte Analytiker?	Führt eine Veränderung des Informationsverhaltens zu einer Veränderung der Effizienz der Analyse?	Gibt es spezifische Wechselwirkungen zwischen Persönlichkeit und Lernverhalten mit Blick auf die Effizienz der Analyse?
Hypothesenformulierung, exemplarisch	Analytiker, die sehr viele oder sehr wenige Informationen beschaffen, erzielen eine geringere Effizienz als Analytiker, die eine mittlere Menge von Informationen beschaffen.	Wenn ein Analytiker seine Informationsbeschaffung stark über sein mittleres Niveau steigert oder darunter senkt, wird er eine geringere Effizienz erzielen als bei seinem mittleren Beschaffungsniveau.	Wenn ein erfahrener Analytiker seine Informationsbeschaffung steigert, ergibt sich eine anderer Effizienzwirkung als bei einem unerfahrenen Analytiker.
Verwendete Meßwerte	Durchschnittliche Meßwerte aus den vier experimentellen Fällen.	Ipsative Meßwerte: Differenzen von fallweisen Meßwerten und individuellen Mittelwerten der vier experimentellen Fällen.	Fallweise ermittelte Meßwerte eines einzelnen Informationsprozesses. Zur Bestimmung der Wechselwirkungen werden Interaktionsvariablen gebildet.
Anwendungsbezogene Tests, exemplarisch	Vergleich von Analytikern, die unterschiedliche Präsentationsformen erhalten: Steigern baumstrukturartige Darstellungen die Effizienz der Analyse?	Vergleich aufeinander folgender Fälle: Bewirken die zentralen Feedbackveranstaltungen und das selbständige, aktive Analysieren einen Lernfortschritt?	Interaktion von Präsentationsform und Lernfortschritt: Er gibt sich bei Baumstrukturen ein anderes Lernverhalten als bei Tabellen und Staffeln?

Die Differenzierung nach interpersoneller und intertemporaler Analyse bedeutet nicht nur eine Verfeinerung des Testdesigns. Sie ist auch von beachtlicher praktischer Bedeutung. Wenn sich zeigen sollte, daß gute Analytiker mehr Informationen beschaffen als schlechte, dann ist aus diesem Befund eine andere Konsequenz zu ziehen als aus dem Ergebnis, daß eine Steigerung der individuellen Informationsbeschaffung zu einer höheren Effizienz der Analyse führt. Im ersten Fall könnte man Personen suchen, die mehr Informationen beschaffen, im zweiten Fall könnte es sich anbieten, das Nachfrageverhalten bereits vorhandener Personen zu vitalisieren.⁷⁷⁾

b) Differenzierung nach dem Inhalt der Fragestellung

Die Bilanzanalyse richtet sich auf zwei große Analysefelder: die Analyse der Erfolgslage und die Analyse der Vermögens- und Finanzlage. Bei der Ableitung unserer Hypothesen haben wir deutlich gemacht, daß bei einer effizienten Analyse die Erfolgslage zuerst bearbeitet werden sollte. In den Schulungsveranstaltungen wird besonderer Wert auf die Analyse der Erfolgslage gelegt. Es könnte daher die Gefahr einer "self-fulfilling prophecy" bestehen.

Um dieser Gefahr entgegenzutreten, wird bei der Effizienzmessung Wert auf "Chancengleichheit" gelegt. Für die Erfolgslage und die Vermögens- und Finanzlage werden gleich viele Urteilsbereiche vorgesehen. Die Gewichtung dieser Bereiche wird nicht nach unserem subjektiven Maßstab, sondern nach externen Normen festgelegt. Maßgeblich für die Gewichtung der Analysefelder sind die statistische Diskriminationsleistung, die fallspezifischen Ausprägungen und die Behandlung der Urteilsbereiche in den bilanzanalytischen Monographien. Da sich die bilanzanalytischen Lehrbücher nachgewiesenermaßen intensiver mit der Vermögens- und Finanzlage befassen, wird dieses Analysefeld sogar begünstigt (vgl. Kapitel 3.2.3.1.).

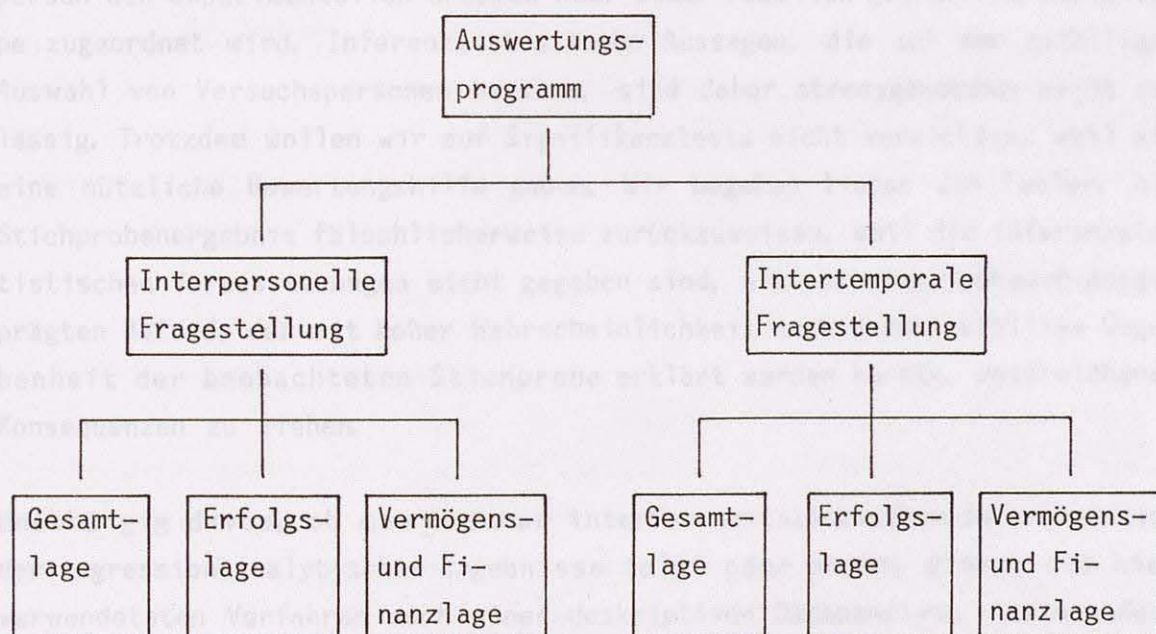
Gleichwohl stellt sich die Frage nach einer differenzierten Effizienzkontrolle der Experimente. Wir haben deshalb sämtliche Dimensionen des Informationsverhaltens und den zentralen Effizienzmeßwert "Treffsicherheit der Analyse" nach den beiden Analysefeldern Erfolgslage und Vermögens- und Fi-

nanzlage getrennt operationalisiert. So kann festgestellt werden, bei **wel-**
chen Inhalten Unterschiede bestehen und ob positive Effekte bei der Analyse
der Erfolgslage durch negative bei der Vermögens- und Finanzlage konterka-
riert werden. Die Differenzierung nach Analysemodulen erlaubt überdies
schärfere Aussagen über die funktionalen Beziehungen zu treffen. Wenn die
Baustein-Hypothese zutrifft, daß eine größere Beschaffungsmenge zu einer
höheren Effizienz führt, dann sollte sich dieser Zusammenhang besonders
deutlich zeigen, wenn man ihn nicht auf die Gesamtmenge sondern auf Teil-
mengen von Informationen bezieht.

c) Zusammenfassung des Auswertungsprogramms

Kombiniert man die uns interessierenden Ausprägungen der Auswertungskrite-
rien, so ergibt sich das in Abbildung 3.18. dargestellte Auswertungspro-
gramm. Die Ergebnisse dieser Tests werden im Befundteil dargestellt. Zu-
sätzlich werden einige Interaktionstests durchgeführt und Ergänzungsbefunde
vorgetragen, die aus vertiefenden Analysen resultieren.

Abbildung 4.18.: Zusammenfassende Darstellung des Auswertungsprogramms



4.2.7.2. Die Prüfung der Verfahrensvoraussetzungen

Bevor dieses Auswertungsprogramm realisiert werden kann, ist ein letzter Prüfschritt erforderlich: die Untersuchung der Voraussetzungen des hier verwendeten regressionsanalytischen Auswertungsverfahrens. Damit wird die vierte Validitätsart von Cook und Campbell geprüft: die "statistical conclusion validity".⁷⁸⁾ Die wesentlichen Voraussetzungen einer interferenzstatistisch eingesetzten Regressionsanalyse sind: zufällig gezogene Beobachtungen, intervallskalierte Skalen, normalverteilte Residuen, Homoskedastizität und das Fehlen von Multikollinearität.⁷⁹⁾ Außerdem sollte der Stichprobenumfang groß sein, um einen praktisch bedeutsamen Effekt mit einer hinreichenden Wahrscheinlichkeit aufzudecken (Kontrolle des sog. "Fehlers zweiter Art").

a) Zufallsstichprobe

Unsere Untersuchungsstichprobe ist keine zufällige, sondern eine systematisch gezogene Stichprobe. Als Versuchspersonen werden nur solche Studenten verwendet, die den Eingangstest bestehen. Innerhalb dieser Grundgesamtheit entscheidet die Leistung des Eingangstests darüber, ob eine Versuchsperson den experimentellen Gruppen oder einer räumlich getrennten Restgruppe zugeordnet wird. Inferenzstatistische Aussagen, die auf der zufälligen Auswahl von Versuchspersonen beruhen, sind daher strenggenommen nicht zulässig. Trotzdem wollen wir auf Signifikanztests nicht verzichten, weil sie eine nützliche Bewertungshilfe geben. Wir begehen lieber den Fehler, ein Stichprobenergebnis fälschlicherweise zurückzuweisen, weil die inferenzstatistischen Voraussetzungen nicht gegeben sind, als aus einem schwach ausgeprägten Befund, der mit hoher Wahrscheinlichkeit auch durch zufällige Gegebenheit der beobachteten Stichprobe erklärt werden könnte, weitreichende Konsequenzen zu ziehen.

Unabhängig davon, ob man unserer interferenzstatistischen Interpretation der regressionsanalytischen Ergebnisse folgt oder nicht, dienen die hier verwendeten Verfahren auch einer deskriptiven Datenanalyse. Insbesondere wird die Differenzierung der Analyse nach durchschnittlichen und ipsativen

Meßwerten und unterschiedlichen Inhalten vorgenommen, um die spezifischen Zusammenhänge möglichst klar herauszuarbeiten.

b) Skalenniveau

Alle Skalen zum Informationsverhalten besitzen (wenigstens) Intervallskalenniveau: Bei der Informationsbeschaffung, Kennzahlenbildung und Übertragung und Speicherung werden Aktivitäten gezählt. Es handelt sich um Skalen mit einem absoluten Nullpunkt und äquidistanten Abständen. Die Anteilsskalen zur Komplexität der Vergleichstechnik und zum Aggregationsgrad stellen ebenfalls Intervallskalen dar. Auch die von Petersen konzipierten Meßwerte zur Sequenz der Verrichtungen und Objekte können als intervallskaliert angesehen werden. Die Rahmenvariablen Präsentationsform und Serie werden durch Dummy-Variablen operationalisiert. Die abhängige Variable "Treffsicherheit der Analyse" stellt eine gewichtete Summe aus verschiedenen Komponenten dar. Die Anforderungen an das Skalenniveau sind somit erfüllt.

c) Fehlen von Multikollinearität

Opp und Schmidt empfehlen, daß die **multiple** Korrelation einer unabhängigen Variablen mit allen anderen unabhängigen Variablen nicht größer als 0,60 sein sollte.⁸⁰⁾ Die Werte unserer Studie liegen im allgemeinen deutlich darunter, so daß mögliche Verzerrungen der Ergebnisse durch Multikollinearitätseinflüsse toleriert werden können.⁸¹⁾

d) Normalverteilte Residuen

Die Normalverteilung der Residuen wird durch optische Inspektion verschiedener Scatterdiagramme und den Kolmogorov-Smirnov-Test überprüft. Im allgemeinen können keine nennenswerten Abweichungen von der Normalverteilung festgestellt werden. Abbildung 3.19. zeigt exemplarisch einen solchen Test (interpersonelle Regression, 103 gemittelte Meßwerte). Die beobachteten

Residuen weichen nur wenig von der geforderten Geraden, die sich bei perfekter Übereinstimmung mit den erwarteten Residuen ergeben müßte, ab. Das Signifikanzniveau des Kolmogorov-Smirnov-Tests bestätigt diese Aussage.⁸²⁾

e) Homoskedastizität

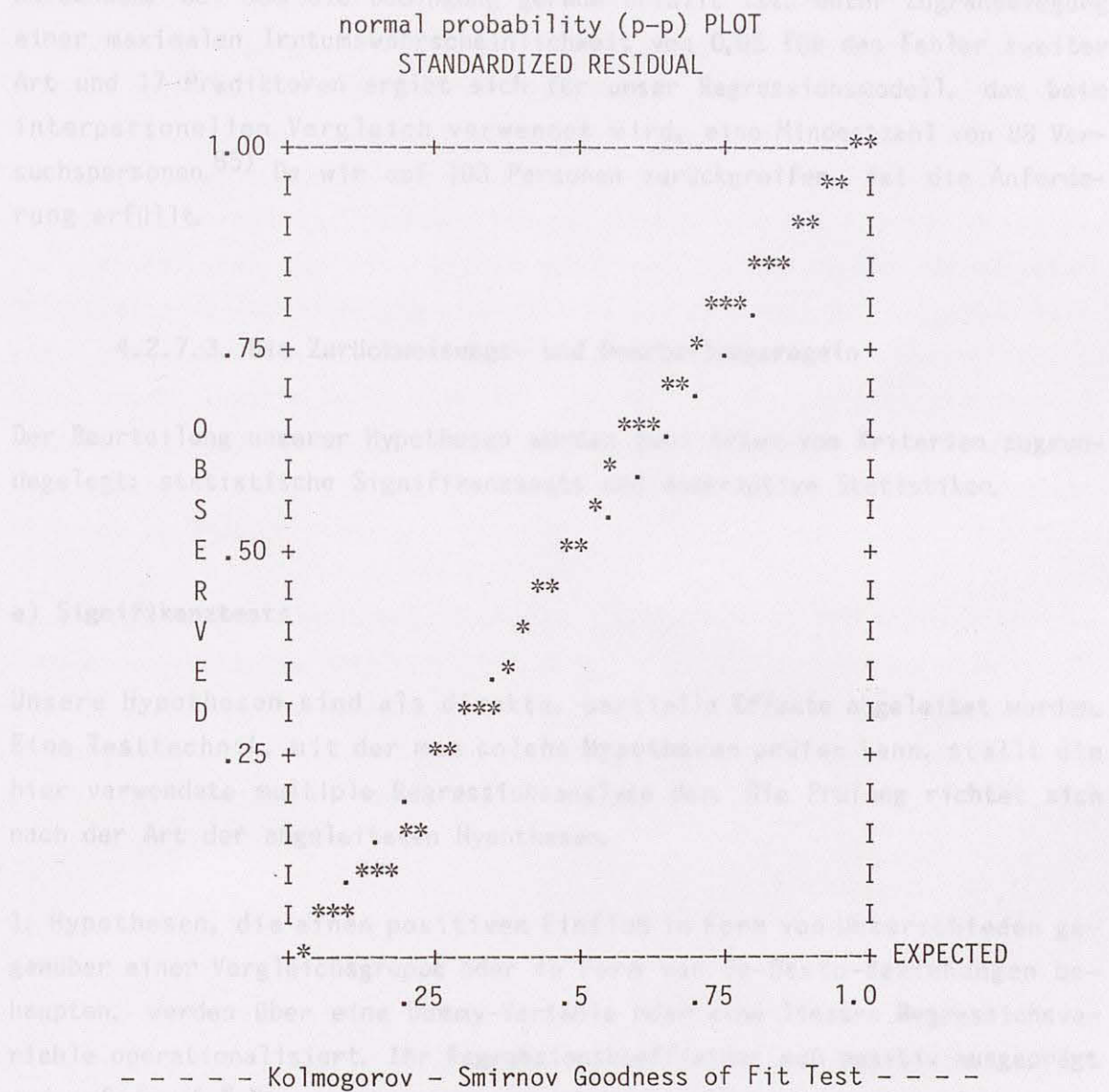
Zur Prüfung der Homoskedastizität werden zwei Varianzanalysen mit Meßwiederholungen gerechnet. In der ersten Analyse werden die "Baum"-Gruppen der Serien Gießen 82 und Kiel 82 verglichen mit den drei Versuchsgruppen der Serie Kiel 84. In der zweiten Analyse werden die beiden "Baum"-Gruppen der Serien Gießen 82 und Kiel 82 mit den "Tabellen"-Gruppen dieser Serien verglichen. In keiner der beiden Analysen kann die Homoskedastizitätsannahme zurückgewiesen werden.⁸³⁾

f) Kontrolle der Teststärke

Zu unseren gerichteten wissenschaftlichen Hypothesen wird eine konvers ausgeprägte statistische Nullhypothese formuliert, die durch einen geeigneten Test überprüft wird. Für den sog. "Fehler erster Art", die fälschliche Verwerfung der statistischen Nullhypothese, wird das gängige 5%-Niveau festgelegt (siehe nachfolgendes Kapitel). Wichtig ist, daß auch man den sog. "Fehler zweiter Art", die fälschliche Verwerfung der Alternativhypothese, kontrolliert. Andernfalls hat man keine Möglichkeit, aus der Annahme der Nullhypothese eine Falsifikation der wissenschaftlichen Hypothese abzuleiten. Da wir aber bestrebt sind, unsere Hypothesen einem entscheidbaren Falsifikationsversuch auszusetzen, muß auch der Fehler zweiter Art kontrolliert werden.

Wir greifen auf einen Vorschlag von Bredenkamp zurück, vor Durchführung des Tests einen praktisch bedeutsamen Effekt festzulegen, der durch das Prüfverfahren mit einer hinreichend großen Wahrscheinlichkeit aufgedeckt wird.⁸⁴⁾ Wir fordern, daß ein Regressionsmodell 25 v. H. der Varianz erklärt. Wenn in der Population ein solcher oder ein größerer Effekt besteht, dann soll er mit einer Wahrscheinlichkeit von mindestens 0,95 zu einer

Abbildung 4.19: Test auf Normalverteilung der Residuen (Interpersonelle Analyse, basierend auf 103 durchschnittlichen Meßwerten)



ZRESID Standardisierte Residuen

Cases	MAX(Abs diff)	MAX(+ diff)	MAX(- diff)
103	0.0500	0.0500	-0.0301
K-S Z	2-tailed P		
0.507	0.959		

signifikanten Verwerfung der Nullhypothese führen.

Zur Beurteilung dieser Anforderung ist der kritische Stichprobenumfang zu berechnen, bei dem die Bedingung gerade erfüllt ist. Unter Zugrundelegung einer maximalen Irrtumswahrscheinlichkeit von 0,05 für den Fehler zweiter Art und 17 Prädiktoren ergibt sich für unser Regressionsmodell, das beim interpersonellen Vergleich verwendet wird, eine Mindestzahl von 88 Versuchspersonen.⁸⁵⁾ Da wir auf 103 Personen zurückgreifen, ist die Anforderung erfüllt.

4.2.7.3. Die Zurückweisungs- und Beurteilungsregeln

Der Beurteilung unserer Hypothesen werden zwei Arten von Kriterien zugrundegelegt: statistische Signifikanztests und deskriptive Statistiken.

a) Signifikanztests

Unsere Hypothesen sind als direkte, partielle Effekte abgeleitet worden. Eine Testtechnik, mit der man solche Hypothesen prüfen kann, stellt die hier verwendete multiple Regressionsanalyse dar. Die Prüfung richtet sich nach der Art der abgeleiteten Hypothesen.

1. Hypothesen, die einen **positiven Einfluß** in Form von Unterschieden gegenüber einer Vergleichsgruppe oder in Form von Je-Desto-Beziehungen behaupten, werden über eine Dummy-Variable oder eine lineare Regressionsvariable operationalisiert. Ihr Regressionskoeffizient muß positiv ausgeprägt und auf dem 5%-Niveau einseitig getestet signifikant sein.

2. Hypothesen, die einen **umgekehrt u-förmigen Verlauf** behaupten, werden durch eine lineare und eine quadratische Regressionsvariable operationalisiert. Dabei wird die ursprüngliche Variable zentriert und die quadratische Variable nach der bei Laroche beschriebenen Formel transformiert, um orthogonale Schätzer für beide Größen zu erhalten.⁸⁶⁾ Eine umgekehrt u-förmige Beziehung setzt voraus, daß der Regressionskoeffizient des quadratischen

Terms negativ ausgeprägt ist.⁸⁷⁾ Wir legen auch bei diesem Koeffizienten einen einseitigen Test mit einem Signifikanzniveau von 5 v. H. zugrunde.

Zum **Ausweis der Signifikanzniveaus** bei den Befundtabellen ist eine klärende Anmerkung erforderlich. Alle Signifikanzangaben für die Regressionskoeffizienten beziehen sich auf **zweiseitige Tests**. Bei den Hauptbefunden werden die Niveaus auf zwei Kommastellen gerundet angegeben. Bei den Detailbefunden wird vermerkt, ob der Koeffizient auf dem 10, 5, oder 1-%-Niveau signifikant ausgeprägt ist. Maßgeblich für diese Form des Ausweises ist das Ziel, einem Leser, der mit **anderen** Hypothesen als wir an den Befund herantritt, eine vollständige und einheitliche Beurteilungsbasis zu geben. Vollständigkeit bedeutet, daß wir auch bei den Regressionskoeffizienten, zu denen unsere Theorie keine Hypothese liefert, ein Signifikanzniveau ausweisen, um eine Beurteilungshilfe zu geben. Einheitlich bedeutet, daß wir alle Signifikanzniveaus auf dieselbe Art von Test beziehen, um die Angaben vergleichen zu können.

Da wir gerichtete Hypothesen vorlegen, testen wir einseitig. Wenn der Befund einen richtungskonträren Verlauf zur statistischen Nullhypothese aufweist, dann wird die Hypothese bis zu einem zweiseitig ausgewiesenen Signifikanzniveau von 10 v. H. verworfen, weil dies einem einseitigen Test mit einem Signifikanzniveau von 5 v. H. entspricht.

b) Deskriptive Statistiken

Über die Hypothesenprüfung hinaus interessiert die Stärke der empirischen Beziehungen, die "praktische", "operative" oder "wissenschaftliche" Signifikanz.⁸⁸⁾ Geeignete Maßzahlen sind in unserem Fall die erklärte Varianz des Regressionsmodells und die standardisierten¹⁵²⁾ Regressionskoeffizienten der einzelnen Variablen. Letztere geben an, in welche Richtung und um wieviele Standardabweichungen sich eine abhängige Variable y verändert, wenn sich nur eine bestimmte unabhängige Variable x_i um eine Standardabweichung verändert. Diese Größen werden in der Literatur auch als "Pfadkoeffizienten" bezeichnet, da sie den **direkten**, partiellen Einfluß messen.⁸⁹⁾ Im Rahmen der Pfadanalyse, die aus mehreren hintereinandergeschalteten Regres-

4.2. Methodik und Durchführung der empirischen Untersuchung Seite - 378 -
sionsanalysen besteht, werden darüber hinaus die **indirekten** und **totalen** Effekte bestimmt.⁹⁰⁾

4.2.1. Die Interpersonelle Analyse

Totale Effekte sind für die Beurteilung der untersuchten Variablen von großer Bedeutung, weil bei alleiniger Betrachtung der direkten Effekte die Gefahr eines **Partialisierungsfehlschlusses** besteht. Hierzu schreiben Hummell und Ziegler: "Nimmt man ... Partialkorrelationen bei 'Konstanz' aller anderen explikativen Variablen als Maß für kausale Relevanz, gelangt man u. U. zu dem absurden Ergebnis, daß überhaupt keine Variable relevant ist. Tatsächlich sind jedoch nur die Effekte einer jeden Variablen mit denen aller anderen konfundiert."⁹¹⁾ Eine alleinige Betrachtung direkter Effekte ist auch aus der Anwendungsperspektive problematisch: Angenommen, eine neuartige Abfolge der Bearbeitung von Analysefeldern würde die Effizienz nur aufgrund indirekter Effekte beachtlich steigern. Bei alleiniger Beurteilung aufgrund des direkten Effekts würde man ein entsprechendes bilanzanalytisches Training als unzumutbar ansehen. Bei angemessener Würdigung des gesamten Variablensystems käme man zu dem gegenteiligen Schluß. Auch **indirekte Effekte** sind wichtig: Durch eine differenzierte Analyse möglicher Saldierungen kann man Anhaltspunkte für eine gezielte Verbesserung der Maßnahme gewinnen.

4.3. DIE ERGEBNISSE DER EMPIRISCHEN UNTERSUCHUNG

4.3.1. Die interpersonelle Analyse

4.3.1.1. Interpersonelle Effizienzunterschiede: Der Vergleich guter und schlechter Analytiker

4.3.1.1.1. Der Ausgangsbefund: Sind alle Effizienzhypothesen zum Informationsverhalten falsifiziert?

Tabelle 4.6. zeigt unseren Befund zur Treffsicherheit der Gesamtlage. Differenziert nach hypothesenbezogenen und datenanalytischen Aussagen ist folgendes festzustellen:

a) Hypothesenbezogene Aussagen:

1. Die Präsentation der verdichteten Informationen in einer **Baumstruktur** führt zwar zu einer höheren Effizienz als die tabellarische Darstellung, aber dieser direkte Einfluß ist nicht signifikant.¹⁾
2. Der behauptete **positive Einfluß der Erfahrung** wird durch den Befund **bestätigt**: Die stärker vorgebildeten Gießener Examenskandidaten analysieren signifikant besser als ihre Kommilitonen aus beiden Kieler Versuchspopulationen.
3. **Alle Hypothesen zum Informationsverhalten**, die einen umgekehrt u-förmigen Verlauf behaupten, werden durch den Befund **widerlegt**: Die Regressionskoeffizienten der quadratischen Terme sind zwar bei fünf der sieben Dimensionen negativ ausgeprägt, aber in keinem Fall wird das geforderte Signifikanzniveau erreicht. Lediglich beim Aggregationsgrad kommt die Irrtumswahrscheinlichkeit in die Nähe des geforderten Werts.²⁾

Tabelle 4.6.: Interpersonelle Analyse – Treffsicherheit der Gesamtlage
(Basis: Je 103 gemittelte Meßwerte)

Explikative Variablen	Hypo- these	Regressionskoeffizienten und Signifikanzniveaus			
		Linearer Term		Quadr. Term	
		Beta	Sign.	Beta	Sign.
Rahmenvariablen					
Baum vs. Tabelle	+	0,14	0,22	- entfällt -	
Gießen vs. Kiel 82	+	0,21	0,08	- entfällt -	
Gießen vs. Kiel 84 ^{&}	+	0,31	0,05	- entfällt -	
Strategische Dimensionen					
Objekt-Gliederung	n	-0,01	0,93	-0,10	0,32
Verrichtungs-Gliederung	n	0,03	0,83	-0,10	0,30
Taktische Dimensionen					
Vergleichstechnik	n	-0,21	0,08	-0,03	0,80
Aggregationsgrad	n	0,24	0,02	-0,14	0,16
Operative Dimensionen					
Beschaffung	n	0,16	0,13	-0,10	0,31
Übertragung	n	-0,21	0,07	0,09	0,40
Kennzahlenbildung	n	0,26	0,02	0,11	0,25

Gesamtmodell: Erklärte Varianz: 33,5 % Signifikanz: $p = 0,0028$

&: Gemeinsamer Effekt von Vorbildung und Präsentationsform

+: Positiver Einfluß n: umgekehrt u-förmiger Zusammenhang

b) Datenanalytische Aussagen:

Daß die geforderten Verlaufsformen nicht bestätigt werden, bedeutet nicht, daß das Informationsverhalten für die Effizienz unbeachtlich ist. Auch bei Konstanthalten der Erfahrung und der Präsentationsform ergeben sich deutlich ausgesprägte, signifikante Effekte bei den **linearen** Regressionsvariablen. Wir orientieren uns bei der Einschätzung dieser Effekte an den zweiseitigen Signifikanzniveaus, da wir zu diesen Effekten keine Hypothesen formuliert haben.³⁾

1. Die beiden **strategischen Dimensionen** zeigen auch beim linearen Term keine signifikant von Null verschiedenen Regressionskoeffizienten. Dies wirft die Frage auf, ob diese Variablen der Prozeßorganisation auf indirekten Pfaden die Effizienz der Analyse beeinflussen.
2. Erstaunlich ist, daß die **Komplexität der Vergleichstechnik** einen tendenziell signifikant **negativen Einfluß** ausübt, während das **Aggregationsniveau** eine signifikant **positive Wirkung** zeigt. Gute Analytiker scheinen sich von schlechten dadurch zu unterscheiden, daß sie sich auf jeweils ein Beurteilungsobjekt konzentrieren und dieses vornehmlich mittels aggregierter Schlüsselinformationen untersuchen.
3. Bei den **operativen Dimensionen** ist auch der lineare Effekt der Informationsbeschaffung vergleichsweise schwach ausgeprägt. Auffällig ist der Gegensatz zwischen dem tendenziell signifikant **negativen Einfluß der Übertragung** und dem signifikant **positiven Einfluß der Kennzahlenbildung**. Gute Analytiker nutzen ihre begrenzte Kapazität zur Informationsverdichtung. Schlechte Analytiker verbrauchen viel Zeit für das Übertragen zahlreicher Detailinformationen. Möglicherweise handelt es sich um ungeübte Analytiker, die sich erst einmal einen Überblick verschaffen wollen, indem sie zahlreiche unverarbeitete Daten auf den Analysebogen übertragen.

So verlockend diese Spekulationen, die sich an die Datenanalyse anschließen, auch sein mögen, es ist nicht unproblematisch, aufgrund des vorliegenden Ausgangsbefunds bereits alle umgekehrt u-förmigen Verlaufshypothesen zu

verwerfen und ad hoc Alternativhypothesen aufzustellen. Es ist vielmehr zweckmäßig, Anschlußtests durchzuführen.

Drei Fragen sind zu beantworten:

1. Beruht die schwache Ausprägung der Regressionskoeffizienten der quadratischen Terme auf einer unzureichenden Variation der unabhängigen Variablen: Haben wir nur einen bestimmten Ast der jeweiligen Kurven beobachten können?
2. Sind die festgestellten Beziehungen genereller Natur, oder ergeben sich bei den einzelnen Fällen unterschiedliche Verläufe? Treten die negativen Effekte vornehmlich bei ungeübten Analytikern auf?
3. Gelten die Effekte auch, wenn man Informationsverhalten und Effizienz nach den Analysefeldern Erfolgslage und Vermögens- und Finanzlage differenziert?

4.3.1.1.2. Prüfung der Variation des Informationsverhaltens: keine Einwände gegen die Tests

Tabelle 4.7. belegt, daß die Dimensionen des Informationsverhaltens über einen beachtlichen Bereich streuen und in mehreren Fällen die theoretisch möglichen Grenzwerte erreichen. Dabei ist zu berücksichtigen, daß die rechnerisch bestimmbaren Extremwerte teilweise sachlich nicht mehr sinnvoll sind: Die minimale Informationsbeschaffungsmenge von Null impliziert, daß der Analytiker keine sachlich begründete bilanzanalytische Aussage treffen kann. Wenn die Vergleichstechnik den Extremwert von 100,0 erreicht, zieht der Analytiker nur Informationen über das Vergleichsunternehmen heran. Da sich nur ein Drittel der angebotenen Informationen auf das Vergleichsunternehmen beziehen und solche Informationen nur für das vorherige Jahr präsentiert werden, verkörpert das empirische Maximum von 38,2 v. H. einen durchaus hohen Wert. **Insgesamt gesehen widerlegt der Befund den Einwand, daß nur ein unzureichender Teil der Variationsbreite der unabhängigen Variablen beobachtet werden konnte.**

Tabelle 4.7.: Variationsbereiche und Mittelwerte der Dimensionen des Informationsverhaltens (Basis: 412 fallweise Meßwerte)

Verhaltensdimensionen	Theoretisches Minimum	Empirisches Minimum	Arithmetisches Mittel	Standardabweichung	Empirisches Maximum	Theoretisches Maximum
Strategische Dimensionen						
Objekt-Gliederung	-100,0	-100,0	32,3	56,8	100,0	100,0
Verrichtungs-Gliederung	-100,0	-46,5	33,6	26,0	100,0	100,0
Taktische Dimensionen						
Vergleichstechnik	0,0	0,0	16,2	12,2	38,2	100,0
Aggregationsgrad	0,0	3,7	59,9	19,8	100,0	100,0
Operative Dimensionen						
Beschaffung	0,0	20,0	90,4	32,0	164,0	390,0
Übertragung	0,0	0,0	38,5	37,3	174,0	unbegrenzt
Kennzahlenbildung	0,0	0,0	20,0	11,7	78,0	unbegrenzt

4.3.1.1.3. Erste Differenzierung: Falleinfluß und Lernfortschritt

Die differenzierende Auswertung nach Analysefällen zeigt folgendes Ergebnis (vgl. Tabelle 4.8.):

1. Die Präsentation der Daten in **Baumstrukturen** bewirkt bei allen vier Versuchen eine **Effizienzsteigerung**. Wenn der Effekt auch in keinem Fall signifikant ist, er ist doch stets vorhanden. Es handelt sich also um einen zwar schwachen, aber zuverlässigen Einfluß.

2. Der direkte Einfluß der **Erfahrung** zeigt sich nur in zwei Fällen, dort aber besonders stark. Die Detailanalyse offenbart, daß die Kandidaten der Serie Kiel 84 von einem besonders geringen Niveau aus starten, aber durch den Einsatz der Baumstruktur während der Schulungsveranstaltungen einen besonders großen Lernfortschritt erleben.⁴⁾ Der große Unterschied zwischen den Gießener Versuchspersonen und den Analytikern der Serie Kiel 84 könnte darauf beruhen, daß der vierte Analysefall, die Pegulan-Werke AG, eine vergleichsweise komplexe Unternehmenssanierung mit einer großangelegten "sale-and-lease-back"-Transaktion darstellt. Bei diesem Fall konnten die Gießener Teilnehmer möglicherweise ein größeres bilanzbezogenes Vorwissen ausspielen.

3. Die **strategischen Dimensionen** zur Prozeßgliederung weisen auch bei den einzelnen Fällen keine umgekehrt u-förmige Beziehung zur Effizienz der Analyse auf. Unsere Überlegungen zur modulweisen Bearbeitung und Integration von Informationen, die aus unterschiedlichen Rechenwerken stammen, sind mit Blick auf den direkten Effekt als falsifiziert anzusehen. Auch die Regressionskoeffizienten der linearen Terme sind im allgemeinen schwach ausgeprägt. Daß sich bei acht getesteten Koeffizienten der linearen Terme in einem Fall ein tendenziell signifikanter Effekt ergibt, sollte angesichts der wechselnden Vorzeichen der schwachen übrigen sieben Effekte nicht überbewertet werden. Zu prüfen bleibt, ob ein starker indirekter Effekt besteht.

Tabelle 4.8.: Interpersonelle Analyse – Treffsicherheit der Gesamtlage differenziert nach Analysefällen (Basis: Je 103 fallweise Meßwerte)

Explikative Variablen		Regressionskoeffizienten bei den Analysefällen			
		1	2	3	4
Rahmenvariablen					
Baum vs. Tabelle		0,12	0,16	0,10	0,10
Gießen vs. Kiel 82		0,09	-0,02	0,14	0,35**
Gießen vs. Kiel 84&		0,30*	0,03	0,01	0,20
Strategische Dimensionen					
Objekt-Gliederung	Lin.	-0,04	-0,06	-0,04	-0,04
	Quad.	0,02	-0,02	-0,03	-0,01
Verrichtungs-Gliederung	Lin.	0,17 ⁺	-0,04	0,00	-0,02
	Quad.	-0,09	-0,03	-0,06	-0,03
Taktische Dimensionen					
Vergleichstechnik	Lin.	-0,12	0,00	-0,11	-0,21
	Quad.	-0,13	-0,11	0,07	0,17
Aggregationsgrad	Lin.	0,15	0,20 ⁺	0,00	0,33**
	Quad.	-0,27**	0,03	-0,17 ⁺	-0,10
Operative Dimensionen					
Beschaffung	Lin.	0,09	0,12	0,24*	0,19
	Quad.	-0,22*	-0,14	-0,06	0,15
Übertragung	Lin.	-0,13	-0,15	-0,28*	-0,35**
	Quad.	0,15	-0,07	0,24*	0,04
Kennzahlenbildung	Lin.	0,20*	0,28**	0,24*	0,11
	Quad.	-0,07	-0,07	0,01	-0,01
Erklärte Varianz:		34,4%**	25,0% ⁺	25,9%*	27,3%*

Signifikanzniveaus: + = 10 %, * = 5 %, ** = 1 % Niveau, zweiseitiger Test

&: Gemeinsamer Effekt von Vorbildung und Präsentationsform

4. Keiner der Regressionskoeffizienten zur **Komplexität der Vergleichstechnik** ist signifikant. Der Koeffizient des linearen Terms ist bei drei von vier Fällen (schwach) negativ ausgeprägt, was dem Gesamteffekt entspricht. Die **Insuffizienzursache** ist eine **mangelhafte Koordination der Informationsaktivitäten**: Wie die Detailanalyse zeigt, gibt es eine Reihe von Analytikern, die Informationen über den Vergleichsfall beschaffen, sie aber bei ihren bilanzanalytischen Urteilen nicht oder nicht sachgerecht nutzen.

5. Beim **Aggregationsgrad** ergibt sich beim ersten und dritten Fall ein signifikant **umgekehrt u-förmiger Verlauf**. Im vierten Fall ist das Vorzeichen hypothesenkonform ausgeprägt. Der Gesamteffekt wird durch den schwach positiven Koeffizienten beim zweiten Fall gedämpft. Wir führen dies auf einen systematischen **Falleinfluß** zurück: Den ersten und dritten Fall kann man lösen, indem man die Summen- und Saldenpositionen der Erfolgsrechnung disaggregiert, weil diese beiden Fälle hier besondere Problembereiche aufweisen. Beim zweiten und vierten Fall ist es zweckmäßiger, auf einer stärker aggregierten Datenebene zu verharren, weil sowohl das aggregierte Bilanzbild als auch das aggregierte Erfolgsbild Probleme signalisieren. Diese Interpretation wird gestützt durch die Tatsache, daß die Koeffizienten der linearen Terme bei beiden Fällen signifikant bzw. tendenziell signifikant positiv ausgeprägt sind.

6. Die Regressionskoeffizienten der **Informationsbeschaffung** weisen ein interessantes Muster auf. Bei den ersten beiden Fällen zeigt sich (zusammengekommen) ein signifikanter, umgekehrt u-förmiger Verlauf. Die linearen Koeffizienten sind unbeachtlich. Ganz anders ist das Bild bei den letzten Fällen. Hier ergibt sich ein signifikant positiver Einfluß auf die Analyse. Das negative Vorzeichen bei den quadratischen Termen verschwindet, wenn man beide Fälle zusammen betrachtet. Dieser Befund entspricht unserer Erwartung, daß sich **negative Effekte** einer selbstinduzierten Informationsüberlastung vor allem bei **ungeübten Analytikern** zeigen sollten.⁵⁾ Er steht im Einklang mit den empirischen Ergebnissen anderer Forscher: Chorba und New stellen fest, daß studentische Versuchspersonen rasch lernen, mit Informationen bewußt umzugehen, wenn sie mit der Aufgabe vertraut werden.⁶⁾

7. Der **negative Effekt der Informationsübertragung** bestätigt sich bei allen vier Fällen. Er zeigt eine im Zeitablauf deutlich zunehmende Tendenz. Ein umgekehrt u-förmiger Verlauf besteht nicht. Schlechte Analytiker zeichnen sich offenkundig dadurch aus, daß sie große Mengen an Informationen erst einmal auf ein Formular übertragen, um sich einen Überblick zu verschaffen. Dieser Vorgang bindet nach unseren Videoaufzeichnungen der Experimente wesentlich mehr Kapazität als die reine Informationsbeschaffung von den Information-Display-Boards.⁷⁾ Er wirkt dysfunktional, weil er zu einer Verdrängung notwendiger Integrationsaktivitäten führt.

8. Bei der **Kennzahlenbildung** zeigt sich in keinem Fall der postulierte umgekehrt u-förmige Verlauf. Wir stellen vielmehr in drei der vier Fälle einen deutlich ausgeprägten, signifikanten **effizienzsteigernden Einfluß** der Kennzahlenbildung fest. Nur bei dem komplexen vierten Fall, bei dem es darauf ankommt, schlecht programmierbare, fallspezifische Verknüpfungen vorzunehmen, fällt der Einfluß schwächer aus.

4.3.1.1.4. Zweite Differenzierung: Effekte unterschiedlicher Analysefelder

Die Differenzierung nach Analysefeldern zeigt zunächst einmal, daß die Beziehungen zwischen Informationsverhalten und Effizienz enger werden: Die erklärte Varianz steigt von 33,5 auf 44,1 bzw. 44,4 v. H. an. Dies ist bei der Vermögens- und Finanzlage vor allem darauf zurückzuführen, daß der positive Einfluß der Kennzahlenbildung auf die Effizienz sehr stark ausgeprägt ist. Bei der Erfolgslage ergibt sich ein starker Einfluß der Nutzung aggregierter Informationen. Im einzelnen sind folgende Aussagen zu treffen (vgl. Tabelle 4.9.):

1. Der Einfluß der **Präsentationsform** zeigt sich nur bei der Erfolgslage. Dies kann damit begründet werden, daß die Baumstruktur hier stärker verzweigt ist und ihre Fähigkeit, Unternehmenskrisen anzuzeigen, empirisch validiert wurde. Es genügt nicht, Informationen in einer bestimmten Form zu präsentieren. Auch die Qualität muß stimmen.

Tabelle 4.9.: Interpersonelle Analyse – Effizienzwirkungen differenziert nach Inhalten (Basis: Je 103 gemittelte Meßwerte)

Explikative Variablen		Regressionskoeffizienten		
		Gesamt- lage	Erfolgs- lage	Vermögens- und Finanzlage
Rahmenvariablen				
Baum vs. Tabelle		0,14	0,14	0,01
Gießen vs. Kiel 82		0,21 ⁺	0,11	0,20 ⁺
Gießen vs. Kiel 84 ^{&}		0,31 [*]	0,21	0,26 [*]
Strategische Dimensionen				
Objekt-Gliederung	Lin.	-0,01	0,01	-0,05
	Quad.	-0,10	0,01	-0,16
Verrichtungs-Gliederung	Lin.	0,03	0,01	0,09
	Quad.	-0,10	-0,08	0,01
Taktische Dimensionen				
Vergleichstechnik	Lin.	-0,21 ⁺	-0,27 [*]	-0,04
	Quad.	-0,03	-0,09	-0,04
Aggregationsgrad	Lin.	0,24 [*]	0,40 ^{**}	0,16
	Quad.	-0,14	-0,16 ⁺	-0,12
Operative Dimensionen				
Beschaffung	Lin.	0,16	0,24 [*]	0,11
	Quad.	-0,10	-0,17 ⁺	-0,19 ⁺
Übertragung	Lin.	-0,21 ⁺	-0,13	-0,16
	Quad.	0,09	0,00	0,13
Kennzahlenbildung	Lin.	0,26 ⁺	0,15	0,46 ^{**}
	Quad.	0,11	-0,03	-0,02
Erklärte Varianz:		33,5% ^{**}	44,4% ^{**}	44,1% ^{**}

Signifikanzniveaus: + = 10 %, * = 5 %, ** = 1 % Niveau, zweiseitiger Test

&: Gemeinsamer Effekt von Vorbildung und Präsentationsform

2. Der Einfluß der **Vorbildung** ist bei der Vermögens- und Finanzlage stärker als bei der Erfolgslage ausgeprägt. Dies ist darauf zurückzuführen, daß die meisten Lehrbücher zur Bilanzanalyse diese Probleme ausführlicher behandeln.

3. Die **strategischen Dimensionen** zur Gliederung des Informationsprozesses zeigen weder bei der Erfolgslage noch bei der Vermögens- und Finanzlage signifikante Einflüsse. Bemerkenswert ist, daß Analytiker, die mit der Erfolgslage beginnen, auch dort nicht signifikant besser analysieren als solche, die mit der Vermögens- und Finanzlage anfangen. Ebenso erstaunlich ist, daß Analytiker, die mit der Vermögens- und Finanzlage beginnen, auch bei diesem Analysefeld nicht signifikant besser abschneiden.

4. Der negative Einfluß der **Vergleichstechnik** gilt nur bei der Erfolgslage. Die Detailanalyse zeigt, daß es bei diesem Analysefeld besonders viele Versuchspersonen die Informationen zum Vergleichsfall nicht sachgerecht nutzen: Im Betriebsvergleich werden trotz unterschiedlicher Betriebsgrößen absolute Erfolgsgrößen statt Renditekennzahlen verwendet. Solche Urteile werden als "nicht sachgerecht" gewertet. Sie gehen nicht in das Effizienzmaß "Treffsicherheit der Lagebeurteilung" ein.

5. Der **Aggregationsgrad** der Informationsaktivitäten zeigt bei beiden Analysefeldern den behaupteten Verlauf. Da er bei einem Analysefeld signifikant ausgeprägt ist und bei dem anderen nicht, ziehen wir zur Bewertung unserer Hypothese eine weitere Differenzierung dieses Befundes nach den einzelnen Analysefällen heran. Sie zeigt bei der Erfolgslage in drei und bei der Vermögens- und Finanzlage in zwei der vier Fälle einen signifikant negativen Regressionskoeffizienten des quadratischen Gliedes. Die gemittelten Regressionskoeffizienten betragen $-0,21$ bzw. $-0,16$. **Wir nehmen daher die Hypothese an**, daß zwischen dem Aggregationsgrad der Informationsaktivitäten und der Effizienz der Analyse eine umgekehrt u-förmige Beziehung bestehe. Bei dieser Aussage ist zu berücksichtigen, daß der Regressionskoeffizient des linearen Terms positiv ausgeprägt ist: Der Abfall der umgekehrt u-förmigen Funktion erfolgt vor allem bei der Erfolgslage relativ spät.

6. Für die **Informationsbeschaffung** ergibt sich bei beiden Analysefeldern ein signifikant ausgeprägter **umgekehrt u-förmiger Verlauf**. Auch diese **Hypothese wird angenommen**.

7. Die **Übertragungsaktivität** zeigt bei beiden Analysefeldern ein negatives Vorzeichen beim linearen Term, aber diese Effekte sind nicht signifikant.

8. Der positive Effekt der **Kennzahlenbildung** stellt sich vor allem bei der Vermögens- und Finanzlage ein. Dies kann darauf zurückgeführt werden, daß man bei der Erfolgslage mit den angebotenen Summen- und Saldenpositionen im Norm- und Zeitvergleich bereits gewisse Aussagen treffen kann: Ein betrieblicher Verlust ist bereits als absolute Größe ein Warnzeichen. Ein negatives Eigenkapital tritt dagegen bei den hier analysierten Aktiengesellschaften nicht auf. Man muß das Eigenkapital wenigstens zu einer zweiten Bilanzposition in Beziehung setzen, um einen sinnvollen Vergleich vornehmen zu können. Generell kann man sagen, daß man sich die Beurteilung der Vermögens- und Finanzlage stark an vertikalen und horizontalen Bilanzstrukturregeln orientiert. D. h. man **berechnet** zuerst eine Kennzahl und nimmt dann einen Vergleich von Kennzahlen vor. Bei der Erfolgslage spielt die Erfolgsquellenzerlegung eine große Rolle. Hier kann man sofort (komplexe) Vergleiche von Niveau und Entwicklung der angebotenen Summen- und Saldenpositionen vornehmen, um das **Muster** der Erfolgsquellenentwicklung zu identifizieren. Dies dürfte ein weiterer Grund sein, warum bei der Erfolgslage die Nutzung aggregierter Daten eine engere Beziehung zur Effizienz aufweist als bei der Vermögens- und Finanzlage.

4.3.1.1.5. Zusammenfassung der direkten Effekte: Informationsverdichtung und Vorbildung sind beachtlich

Die Befunde können auf folgende Kernaussagen reduziert werden:

1. **Maßgeblich für eine gute Analyse ist die Informationsverdichtung.** Ihre Schlüsselrolle zeigt sich im positiven Einfluß der Kennzahlenbildung und in der effizienzsteigernden Wirkung der genutzten Summen- und Saldenpositionen. Dieses Ergebnis bedeutet eine Bestätigung der bilanzanalyti-

schen Literatur, die sich intensiv mit Fragen der Kennzahlenbildung beschäftigt. Es unterstreicht den potentiellen Nutzen von Computern für die Bilanzanalyse. Das Ergebnis ist allerdings in zwei wichtigen Punkten zu qualifizieren: Die angebotenen Informationen müssen eine prognostische Relevanz besitzen, mögliche positive und negative Unternehmensentwicklungen frühzeitig anzuzeigen.⁸⁾ Die Informationsaktivität darf sich nicht auf aggregierte Daten beschränken, sondern muß diese in disaggregierte Daten zerlegen und mit solchen verknüpfen, sonst greift der fallende Ast der Effizienzfunktion.

2. **Symptomatisch für schlechte Analysen sind große Mengen übertragener Informationen:** Der zeitaufwendige Kopiervorgang stiftet offenbar nur einen geringen Erkenntnisfortschritt. Er verdrängt notwendige Integrationsaktivitäten.
3. Ein weiteres Kennzeichen einer schlechten Analyse ist eine **mangelhafte Koordination der Informationsaktivitäten**, namentlich die Beschaffung und Übertragung von Informationen über den Vergleichsfall, die nicht bei der Urteilsbildung berücksichtigt werden.
4. Der Einfluß der **Informationsbeschaffung** ist bei **erfahrenen Analytikern positiv**. Dysfunktionale Informationsüberlastungseffekte nehmen mit zunehmender Erfahrung ab.
5. Die **Ablauforganisation des Informationsprozesses nach Objekten und Verrichtungen hat keinen Einfluß auf die Effizienz** – zumindest keinen direkten. Ob es dennoch eine optimale Organisationsform gibt, die für alle Analytiker die beste aller möglichen darstellt, muß die Analyse der indirekten Effekte zeigen.
6. Gute und schlechte Analytiker unterscheiden sich durch ihre **Vorbildung**. Der durchschnittliche Effekt ist zwar deutlich, aber doch nicht bei jedem zu analysierenden Fall zu erwarten. Es kommt vielmehr darauf an, daß die Fallanforderungen dem Vorwissen entsprechen. Die "Erfahrung" erweist sich auch in unserer Studie als ein vielschichtiges Konstrukt.

7. Bei der **Präsentationsform** zeigt sich nur ein schwacher, aber stets vorhandener Effekt. Er ist bei der Erfolgslage stärker ausgeprägt als bei der Vermögens- und Finanzlage. Dies ist mit stärker verzweigter Baumstruktur, den besser trennenden Erfolgsinformationen und dem Schulungsprogramm zu erklären, das besonderen Wert auf die Erfolgsquellenzerlegung legt. Auch bei diesem Instrument muß die Analyse der indirekten Effekte abgewartet werden.

Bemerkenswert ist, daß nur zwei unserer sieben Hypothesen über umgekehrt u-förmige Beziehungen zwischen Informationsverhalten und Effizienz beibehalten werden können – und auch hier ist die empirische Evidenz nicht sehr stark. Es gibt demnach wohl extreme Verhaltensweisen, die als insuffizient angesehen werden können, aber allgemeingültige Aussagen über ein "optimales" Informationsverhalten sollte man nicht formulieren, zumindest nicht auf Basis der hier vorgelegten Befunde.

Unseres Wissens gibt es nur zwei weitere empirische Studien, in denen eine **nicht**-lineare Beziehung zwischen Informationsverhalten und Effizienz einer betriebswirtschaftlichen Aufgabe getestet wird. Dies sind die Untersuchungen von Shields und Brockhoff. Shields läßt Kosten- und Leistungsberichte (sog. "performance reports") von Managern analysieren. Er ermittelt zwar eine umgekehrt u-förmige Beziehung zwischen Informationsbeschaffung und Urteilsqualität, aber die Koeffizienten sind nicht signifikant. Die erklärte Varianz beträgt nur 1 v. H.⁹⁾ Brockhoff führt Delphi-Prognosen im Computerdiallog mit Wirtschaftspraktikern aus verschiedenen Branchen durch. Seine Funktionsschätzungen zeigen nur schwach gekrümmte Verläufe für den Zusammenhang von Datenbanknutzung und Prognoseleistung.¹⁰⁾

Man sollte daher mit generalisierenden, vereinfachenden Aussagen über "optimales" Informationsverhalten zurückhaltender sein.¹¹⁾ Es gibt einfach zu große inter-individuelle Unterschiede, um "das" für alle Menschen und Aufgaben gültige Verhalten zu postulieren. Sinnvoller, wenn auch weniger aufmerksamkeitsregend, scheint uns eine Suche nach den spezifischen individuellen Insuffizienzursachen des Informationsverhaltens zu sein. Wenn man die schlechte Lösung verhindert, dann kommt man der guten schon ein Stück näher – auch wenn man die beste nicht erreicht.

4.3.1.2. Interpersonelle Unterschiede im Informationsverhalten

4.3.1.2.1. Die Einflüsse der Rahmenvariablen: deutliche Effekte der Vorbildung

Tabelle 4.10. zeigt, welchen Einfluß die Rahmenvariablen auf das Informationsverhalten ausüben:

1. Alle Hypothesen zum Einfluß der Baumstruktur werden falsifiziert.¹²⁾

Dennoch ist die Präsentationsform nicht unbeachtlich. Sie bewirkt eine Veränderung des Analysekonzeptes: Versuchspersonen, die eine Baumstruktur erhalten, analysieren die Erfolgslage signifikant früher. Sie werden ange-regt, mehr Informationen zu beschaffen. Das Niveau der Kennzahlenbildung ist bei Baumstrukturen etwas geringer.

Der Einfluß der Baumstruktur auf die Prozeßgliederung wird durch den Kon-trast "Gießen versus Kiel 84" bestätigt. Der negative Regressionskoef-fizient von $-0,41$ bei der Objektgliederung besagt, daß die Kieler Analyti-ker wesentlich häufiger mit der Erfolgslage beginnen. Dieser Effekt beruht aber nur zum kleineren Teil auf der unterschiedlichen Population, denn die Variable "Gießen vs. Kiel 82" zeigt nur einen Unterschied von $0,06$. Es bleibt eine Differenz von $0,35$ übrig. Sie kann darauf zurückgeführt werden, daß bei der Serie Kiel 84 die Baumstrukturen **zusätzlich** während der zentra-len Schulungsveranstaltungen verwendet werden. In ähnlicher Weise bestätigt der negative Regressionskoeffizient von $-0,24$ den dämpfenden Einfluß des Baumstrukturkonzeptes auf die Kennzahlenbildung.

2. Betrachtet man die Regressionskoeffizienten der Variablen "Gießen versus Kiel 82", so ist festzustellen, daß zwei der fünf Hypothesen zum Einfluß der **Erfahrung** bestätigt werden: Die stärker vorgebildeten Gießener Teilneh-mer weisen eine signifikant komplexere Vergleichstechnik auf. Sie nutzen signifikant mehr disaggregierte Informationen. Die noch stärker ausgepräg-ten Regressionskoeffizienten des Kontrastes "Gießen vs. Kiel 84" bestätigen diesen Befund, denn die Analytiker der Serie Kiel 84 weisen die geringste Vorbildung auf.

Tabelle 4.10.: Interpersonelle Analyse - Einflüsse der Rahmenvariablen auf das Informationsverhalten (Basis: Je 103 gemittelte Meßwerte)

Abhängige Variablen	Explikative Rahmenvariablen				
	Präsentationsform		Erfahrung		Komb. Effekt Gießen vs. Kiel 84
Dimensionen des Informationsverhaltens	Hypo- these	Baum vs. Tabelle	Hypo- these	Gießen vs. Kiel 82	
Strategische Dimensionen Objekt-Gliederung Verrichtungs-Gliederung	P	0,31**	P	-0,06	-0,41**
	P	0,01	P	0,05	0,05
Taktische Dimensionen Vergleichstechnik Aggregationsgrad	+	-0,03	+	0,27**	0,60**
	-	0,09	-	-0,27*	-0,44**
Operative Dimensionen Beschaffung Übertragung Kennzahlenbildung	-	0,19 ⁺	+	-0,10	0,06
	-	0,02	+	0,11	-0,11
	+	-0,14	+	-0,02	-0,24 ⁺

Signifikanzniveaus: + = 10 %, * = 5 %, ** = 1 % Niveau, zweiseitiger Test

P: Gegenstand der Arbeit von Petersen

+: Positiver Zusammenhang

-: Negativer Zusammenhang

In Tabelle 4.11. sind die **Befunde nach Erfolgslage und Vermögens- und Finanzlage differenziert:**

1. Unsere Hypothesen zur Wirkung der **Baumstruktur** sind auch bei einer Differenzierung nach Erfolgslage und Vermögens- und Finanzlage als falsifiziert anzusehen.

Die Datenanalyse weist auf eine wichtige Differenzierung hin: Die Baumstruktur wirkt bei der Erfolgslage anders als bei der Vermögens- und Finanzlage. Bei der **Erfolgslage** führt sie zu einer **Vitalisierung der Informationsbeschaffung** und zu einer stärkeren Nutzung aggregierter Daten. Bei der Vermögens- und Finanzlage gibt es keine derartigen Effekte. Der Befund offenbart eine interessante **Wechselwirkung von Schulung und Präsentationsform**: Nur bei den Versuchsgruppen, bei denen die Präsentationsform des Angebots dem theoretischen Konzept entspricht, werden die Analytiker angeregt, mehr Informationen zu beschaffen. Daß sich beim Vergleich "Gießen vs. Kiel 84" ein geringerer Unterschied ergibt, liegt daran, daß sich die Analytiker der Serie Kiel 84 weitgehend auf Zeitvergleiche beschränken. Betrachtet man nur die für den Zeitvergleich beschafften Informationen, fällt der Effekt deutlich höher aus.

2. Die festgestellten Einflüsse der **Erfahrung** gelten sowohl für die Erfolgslage als auch für die Vermögens- und Finanzlage. Unsere Hypothesen zur Komplexität der Vergleichstechnik und zum Aggregationsgrad werden damit erhärtet und als bestätigt angesehen.

Die Übertragungsmenge ist zwar bei der Vermögenslage signifikant höher ausgeprägt (bei hypothesenbezogenem einseitigen Test), aber die Einflüsse auf die Kennzahlenbildung und die Informationsbeschaffung sind schwach und nicht signifikant. Unsere **generelle** Hypothese, daß Analytiker mit einer größeren Erfahrung ein größeres Aktivitätsniveau aufweisen, wird daher als falsifiziert angesehen.

Tabelle 4.11.: Interpersonelle Analyse - Einflüsse der Rahmenvariablen auf das Informationsverhalten differenziert nach Analysefeldern
(Basis: Je 103 gemittelte Meßwerte)

Abhängige Variablen	Explikative Rahmenvariablen					
	Präsentationsform		Erfahrung		Kombinierter Effekt	
	H. Baum vs. Tabelle	H. Gießen vs. Kiel 82	H. Gießen vs. Kiel 82	H. Gießen vs. Kiel 82	Gießen vs. Kiel 82	Verfin.
Strategische Dimensionen	Ges.	Ges.	Ges.	Ges.	Ges.	Verfin.
	Objekt-Gliederung	Objekt-Gliederung	Objekt-Gliederung	Objekt-Gliederung	Objekt-Gliederung	Objekt-Gliederung
Taktische Dimensionen	Ges.	Ges.	Ges.	Ges.	Ges.	Verfin.
	Verrichtungs-Gliederung	Verrichtungs-Gliederung	Verrichtungs-Gliederung	Verrichtungs-Gliederung	Verrichtungs-Gliederung	Verrichtungs-Gliederung
Operative Dimensionen	Ges.	Ges.	Ges.	Ges.	Ges.	Verfin.
	Beschaffung	Beschaffung	Beschaffung	Beschaffung	Beschaffung	Beschaffung
Kennzahlenbildung	Ges.	Ges.	Ges.	Ges.	Ges.	Verfin.
	Übertragung	Übertragung	Übertragung	Übertragung	Übertragung	Übertragung

Signifikanzniveaus: + = 10 %, * = 5 %, ** = 1 % Niveau, zweiseitiger Test

H.: Hypothese

P: Gegenstand der Arbeit von Petersen

+: Positiver Zusammenhang

-: Negativer Zusammenhang

3. Besonders starke Unterschiede zeigt der Kontrast **"Gießen vs. Kiel 84"**. Hier summieren sich die Einflüsse von Präsentationsform und Population. Die weniger stark vorgebildeten Kieler Probanden haben das Analysekonzept, das ihnen zusätzlich in graphischer Form präsentiert wird, besonders stark und besonders häufig übernommen: Sie beginnen überwiegend mit der Analyse der Erfolgslage, nutzen die aggregierten Daten intensiver und konzentrieren sich auf Zeitvergleiche. Diese Vorgehensweise entspricht den vorgetragenen Musterlösungen. In **"Wort und Bild"** wird ihnen erklärt, wie man die Baumstruktur zur Erfolgsquellenzerlegung systematisch nutzen kann. Dabei wird der Betriebsvergleich nur bei geeigneten Schlüsselpositionen eingesetzt. Die gleichzeitige Verwendung der Präsentationsform beim Vortrag und beim Informationsangebot weist einen beachtlichen Interaktionseffekt auf: Aus dem **"äußeren"** graphischen Modell wird ein **"inneres"**, mental aufgenommenes Analysekonzept.

4.3.1.2.2. Die Dimensionen des Informationsverhaltens: unabhängige Aktionsparameter

Tabelle 4.12. zeigt die Beziehungen zwischen den Dimensionen des Informationsverhaltens. Unsere Interpretation wird nach hypothesenbezogenen und datenanalytischen Aussagen differenziert.

a) Hypothesenbezogene Aussagen

1. Operative Dimensionen: Die Regressionskoeffizienten bestätigen unsere erste Hypothesenvariante. Alle drei Zusammenhänge sind positiv ausgeprägt, davon sind zwei signifikant. Der in der zweiten Hypothesenvariante behauptete Kapazitätskonflikt tritt nur zwischen den Aktivitäten **"Übertragung und Speicherung"** und **"Kennzahlenbildung"** auf. (Vgl. den signifikant negativ ausgeprägten Koeffizienten des quadratischen Gliedes, der einer umgekehrt u-förmigen Beziehung entspricht). Wir vermuten, daß man eine große Menge beschaffter Informationen zwar auf den Analysebogen übertragen kann, aber dann fehlt die Zeit, um sie zu Kennzahlen zu verdichten. Der vergleichsweise stark ausgeprägte signifikant negative Zusammenhang zwischen der Verrichtungsgliederung und der Kennzahlenbildung bestätigt diese Hypothese. Es

Tabelle 4.12.: Interpersonelle Analyse - Die Zusammenhänge zwischen den Dimensionen des Informationsverhaltens (Basis: Je 103 gemittelte Meßwerte)

Abhängige Variablen	Explikative Variablen					
	Strategische Dimensionen		Taktische Dimensionen		Operative Dimensionen	
	Objekt- Gliederung H. Lin. Quad.	Verrichtungs- Gliederung H. Lin. Quad.	Vergleichs- technik H. Lin. Quad.	Aggregations- niveau H. Lin. Quad.	Informations- beschaffung H. Lin. Quad.	Übertragung und Speicherung H. Lin. Quad.
Dimensionen des Informationsverhaltens						
Strategische Dimensionen						
Verrichtungs-Gliederung	P -0,33** 0,06					
Taktische Dimensionen						
Vergleichstechnik	0 0,08 -0,09	0 0,26** 0,12				
Aggregationsgrad	0 0,05 0,24**	0 0,19+ 0,02	+ 0,01 -0,01			
Operative Dimensionen						
Beschaffung	0 -0,02 -0,09	0 0,00 -0,03	0 0,23+ -0,13	0 0,07 -0,02		
Übertragung	0 -0,22* 0,14	0 0,01 0,01	0 0,36** 0,22*	0 0,11 0,04	+ / n 0,28** -0,04	
Kennzahlenbildung	0 0,06 -0,11	0 -0,39** -0,04	0 0,06 0,19*	0 -0,05 -0,11	+ / n 0,14 0,09	+ / n 0,29** -0,17+

Signifikanzniveau: + = 10 %, * = 5 %, ** = 1 % Niveau, zweiseitiger Test

P: Gegenstand der Arbeit von Petersen

+: Positiver Zusammenhang

0: Kein systematischer Zusammenhang

n: Umgekehrt u-förmiger Zusammenhang

handelt sich um Analytiker, die sich vom Informationsangebot oder einem selbst entworfenen **Formular**¹³⁾ leiten lassen. Was beschafft wird, entscheidet das Formular, nicht bereits die von ihnen beurteilten Bereiche. Dies führt dazu, daß man viele Informationen nachfragt, die nicht mehr zu Kennzahlen verdichtet werden. **Die phasentheoremkonforme, formular- statt problemorientierte Vorgehensweise begünstigt einen unvollständigen und unwirtschaftlichen Informationsprozeß.**

2. Taktische Dimensionen: Zwischen der Komplexität der Vergleichstechnik und dem Aggregationsgrad ergibt sich kein umgekehrt u-förmiger Zusammenhang. Dies kann daran liegen, daß beide Dimensionen durch **Anteilsskalen** operationalisiert werden. Unserer Hypothesenableitung liegt ein Kapazitätskonflikt zugrunde, der sich auf das **Niveau** beider Verhaltensdimensionen stützt.

b) Datenanalytische Aussagen

1. Zwischen der Objektgliederung und der Nutzung aggregierter Daten besteht ein u-förmiger Zusammenhang. Dies könnte auf folgender Vorgehensweise beruhen: Analytiker, die extreme Werte auf der Objektgliederungsskala aufweisen, arbeiten eines der beiden Analysefelder vollständig ab, bevor sie sich dem zweiten zuwenden. Sie greifen dabei vergleichsweise stärker auf disaggregierte Informationen zurück, als solche Analytiker, die beide Felder zunächst auf einer hohen Aggregationsebene untersuchen und sich dann den Detailbefunden zuwenden.

2. Wer zuerst die Erfolgslage analysiert, überträgt weniger Informationen. Dies entspricht der empirischen Beobachtung, daß sich manche Analytiker die gezogenen Informationskärtchen zur Erfolgslage in gedachten Baumstrukturen auf ihrem Arbeitsplatz auslegen und nur ihre Urteile auf die Analysebögen schreiben.

3. Bei einer hohen Komplexität der Vergleichstechnik zeigen sich ähnlich dysfunktionale Effekte wie bei der strengen Gliederung nach Verrichtungen: Wer Zeit- und Betriebsvergleiche **anstrebt**, beschafft und überträgt signifi-

kant mehr Informationen, aber er kann sie nicht mehr zu Kennzahlen verdichten, denn zu dieser Aktivität besteht kein signifikanter Zusammenhang. Die bereits festgestellten negativen direkten Effizienzwirkungen der Komplexität der Vergleichstechnik belegen, daß nicht nur die Kennzahlenberechnung, sondern vor allem die daran anknüpfende **Beurteilung** der Daten zu kurz kommt. **Das Bestreben, das Informationsangebot möglichst vollständig auszunutzen, führt zu einer unwirtschaftlichen, selbst induzierten Informationsüberlastung.**

Komplexere Vergleichstechniken sind nicht schematisch, sondern relevanzgesteuert anzuwenden. Die Ursache für die dysfunktionale Wirkung der Vergleichstechnik besteht unseres Erachtens in einer mangelhaften Koordination der Informationsaktivitäten. Daß wir **keinen** Konflikt zwischen Aggregationsgrad und Vergleichstechnik feststellen können, liegt wohl nicht nur an der Meßtechnik, sondern auch daran, daß viele Analytiker **gleichzeitig** in die Breite und die Tiefe gehen wollen. Diesen Plan können sie jedoch nur bei der Beschaffung und Übertragung verwirklichen. Danach stoßen sie an ihre Kapazitätsgrenze. Es ist deshalb wichtig, daß man Analysetechniken vermittelt, die es erlauben, die Menge der möglichen Vergleiche auf sachlich wichtige Informationen zu beschränken. Nur so kann der Analytiker das empirisch nachgewiesene Informationspotential des Betriebsvergleichs nutzen.

4. Nachrichtlich ist auf die signifikant negative Beziehung zwischen der **Gliederung des Prozesses** nach Objekten und Verrichtungen hinzuweisen. Das Verhältnis zwischen diesen Gliederungsprinzipien wird in der Arbeit von Petersen systematisch untersucht. Für uns ist wichtig, daß die beiden Dimensionen nur mäßig miteinander korrelieren, so daß wir sie als zwei unabhängige Variablen in unseren Regressionsgleichungen verwenden dürfen. Die Inspektion der Streupunktdiagramme belegt, daß ein großer Teil der möglichen Fläche abgedeckt wird und das Ergebnis nicht durch Ausreißer zustande kommt. Es handelt sich um weitgehend unabhängige Aktionsparameter.

4.3.1.3. Die Pfadanalyse der Effizienzwirkungen

4.3.1.3.1. Zur Methodik der Auswertungen

Unsere bisherigen Auswertungen richteten sich auf die Prüfung der postulierten direkten Effekte. Nun sollen die indirekten und totalen Effekte bestimmt werden. Damit verlagert sich das Auswertungsziel von der Hypothesenprüfung auf die Datenanalyse. Es geht darum, eine Beziehungsstruktur zu finden, die die wesentlichen Zusammenhänge unserer Stichprobe aufzeigt.

Auch diese Analyse basiert auf bestimmten Prämissen:

- a) Wir gehen von linearen Beziehungen aus.
- b) Wir gehen von einer bestimmten Präzedenzstruktur der Variablen aus.

Zu a) Lineare Beziehungen

Die prognostizierten quadratischen Terme haben sich bei unserer Analyse selten signifikant absichern lassen. Testet man die zusätzliche Varianzerklärung aller quadratischen Terme zusammengekommen, so ergibt sich weder bei der Gesamtlage noch bei den einzelnen Analysefeldern ein signifikanter Zuwachs gegenüber dem linearen Modell. Berechnet man die Pfadmodelle mit linearen und quadratischen Variablen, so zeigt sich bei keiner quadratischen Variablen ein nennenswerter indirekter Effekt. Geht man von der quadratischen zur linearen Pfadanalyse über, so beträgt die mittlere absolute Veränderung der totalen Effekte der linearen Terme nur 0,01. Es ist daher sinnvoll, sich auf das wesentlich übersichtlichere lineare Modell zu beschränken. Um Fehlinterpretationen zu vermeiden, werden die bereits festgestellten **direkten** Effekte quadratischer Terme bei der Befundbeurteilung berücksichtigt.

Zu b) Unterstellte Einflußstruktur

Die ermittelten indirekten und direkten Effekte gelten nur auf Basis einer unterstellten Einflußstruktur. Die nachfolgende Präzedenzmatrix zeigt, von welcher Einflußstruktur wir ausgehen. Eine Eins bedeutet, daß die entsprechende Spaltenvariable die entsprechende Zeilenvariable beeinflusst. Bei den beiden Dimensionen zur Gliederung des Prozesses ist keine allgemeingültige Einflußrichtung bestimmbar. Es ist sowohl der Fall möglich, daß ein Analytiker den Prozeß primär nach Objekten gliedert und die Verrichtungsgliederung nur noch innerhalb der Objekte zum Zuge kommt, als auch der umgekehrte Fall. Deshalb steht an dieser Stelle der Matrix eine Null. Auch bei den Zusammenhängen zwischen den Rahmenvariablen steht eine Null: Die Präsentationsform und die Vorbildung der Analytiker sind Parameter unseres experimentellen Designs, die von uns in einer bestimmten Weise variiert werden.

Abbildung 4.20.: Präzedenzmatrix der untersuchten Variablen

[illegible]

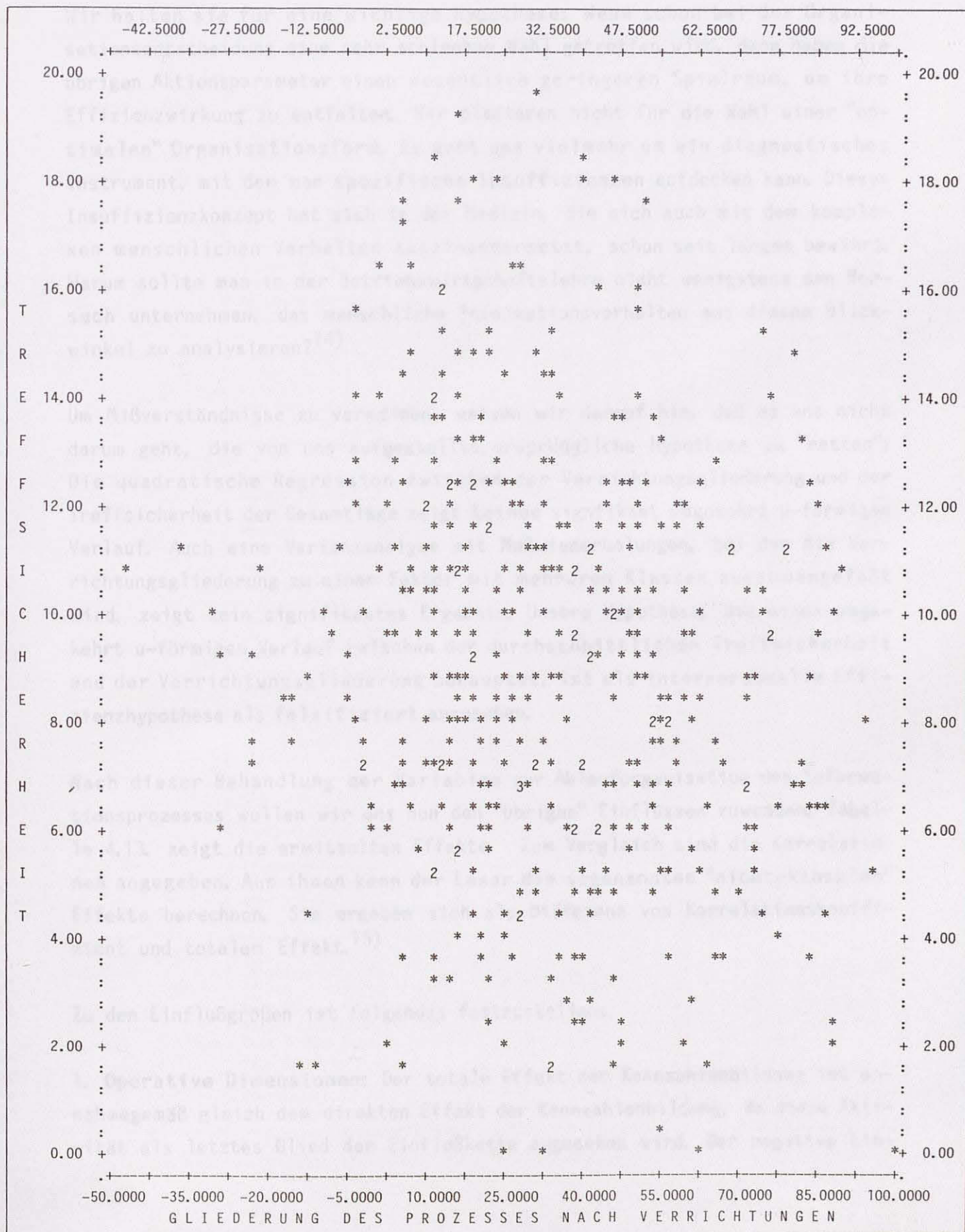
4.3.1.3.2. Treffsicherheit der Gesamtlage: dysfunktionale indirekte Effekte der Vorbildung

Ein zentraler Befund soll vorweggenommen werden: Die Dimensionen zur Ablauforganisation des Informationsprozesses zeigen auch bei Berücksichtigung der indirekten Effekte keinen nennenswerten Einfluß auf die Treffsicherheit der Analyse. Diese Aussage gilt nicht nur für die Treffsicherheit der Gesamtlage, sondern auch für die Erfolgslage und für die Vermögens- und Finanzlage: Wenn unterschiedliche Analytiker den Ablauf eines bilanzanalytischen Informationsprozesses unterschiedlich gliedern, dann hat dies keinen Einfluß auf die durchschnittlich zu erwartende Effizienz der Analyse. **Die Vorstellung eines "one best way of process organization" – wie sie in manchen Phasenvorschlägen anklingt – ist als falsifiziert anzusehen.**

Dennoch ist die Prozeßorganisation nicht völlig unbeachtlich. Zur Illustration des Gemeinten wird der Leser gebeten, das Streupunktdiagramm der Abbildung 4.21. zu betrachten. Es zeigt den Zusammenhang zwischen der Gliederung des Prozesses nach Verrichtungen und der Treffsicherheit der Gesamtlage.

Man kann mit einigem Wohlwollen erkennen, daß die **umhüllende** Kurve des Streupunktdiagramms ein umgekehrtes U beschreibt – wobei sich innerhalb dieser Hülle eine ganz beachtliche Variation entfaltet. In dieser Form mag man sich die Wirkung der Organisation vorstellen: Sie setzt einen **Rahmen**, innerhalb dessen sich eine ganz unterschiedliche Effizienzwirkung aufgrund der "übrigen" Parameter ergeben kann. Die Hypothese sollte daher nicht lauten: Wenn eine Organisation der Form o gewählt wird, dann ist im Durchschnitt mit einer Effizienz der Ausprägung e zu rechnen. Es sollte vielmehr heißen: Wenn eine Organisation der Form o gewählt wird, dann ist **höchstens** mit einer Effizienz der Ausprägung e zu rechnen. In diesem Sinne legt unser Befund folgende Hypothese nahe: **Wenn eine extrem phasenkonforme oder eine extrem phasenkonträre Gliederung der Verrichtungen des bilanzanalytischen Informationsprozesses gewählt wird, dann ist höchstens mit einer mittleren Effizienz zu rechnen.** Dies ist natürlich eine neue **Hypothese**, die mit einem dafür besonders geeigneten Testverfahren an einer neuen Stichprobe geprüft werden sollte.

Abbildung 4.21.: Interpersonelle Analyse - Streupunktdiagramm von Verrichtungs-Gliederung und Treffsicherheit der Gesamtlage (Basis: 412 fallweise Meßwerte)



Wir halten sie für eine wichtige Hypothese: Wenn schon bei der Organisationsentscheidung eine sehr schlechte Wahl getroffen wird, dann haben die übrigen Aktionsparameter einen wesentlich geringeren Spielraum, um ihre Effizienzwirkung zu entfalten. Wir plädieren nicht für die Wahl einer "optimalen" Organisationsform. Es geht uns vielmehr um ein diagnostisches Instrument, mit dem man **spezifische Insuffizienzen** aufdecken kann. Dieses Insuffizienzkonzept hat sich in der Medizin, die sich auch mit dem komplexen menschlichen Verhalten auseinandersetzt, schon seit langem bewährt. Warum sollte man in der Betriebswirtschaftslehre nicht wenigstens den **Versuch** unternehmen, das menschliche Informationsverhalten aus diesem Blickwinkel zu analysieren?¹⁴⁾

Um Mißverständnisse zu vermeiden, weisen wir darauf hin, daß es uns nicht darum geht, die von uns aufgestellte, ursprüngliche Hypothese zu "retten": Die quadratische Regression zwischen der Verrichtungsgliederung und der Treffsicherheit der Gesamtlage zeigt **keinen** signifikant umgekehrt u-förmigen Verlauf. Auch eine Varianzanalyse mit Meßwiederholungen, bei der die Verrichtungsgliederung zu einem Faktor mit mehreren Klassen zusammengefaßt wird, zeigt kein signifikantes Ergebnis. Unsere Hypothese, die einen umgekehrt u-förmigen Verlauf zwischen der durchschnittlichen Treffsicherheit und der Verrichtungsgliederung behauptet, ist als interpersonelle Effizienzhypothese als **falsifiziert** anzusehen.

Nach dieser Behandlung der Variablen zur Ablauforganisation des Informationsprozesses wollen wir uns nun den "übrigen" Einflüssen zuwenden. Tabelle 4.13. zeigt die ermittelten Effekte. Zum Vergleich sind die Korrelationen angegeben. Aus ihnen kann der Leser die sogenannten "nicht-kausal" Effekte berechnen. Sie ergeben sich als Differenz von Korrelationskoeffizient und totalem Effekt.¹⁵⁾

Zu den Einflußgrößen ist folgendes festzustellen:

1. Operative Dimensionen: Der totale Effekt der Kennzahlenbildung ist annahmegemäß gleich dem direkten Effekt der Kennzahlenbildung, da diese Aktivität als letztes Glied der Einflußkette angesehen wird. Der negative Ein-

Tabelle 4.13.: Interpersonelle Analyse – Pfadanalyse zur Treffsicherheit der Gesamtlage (tabellarischer Befund)
(Basis: Je 103 gemittelte Meßwerte)

Abhängige Variablen	Unabhängige Variablen	Erklärte Varianz	Direkter Effekt	Totaler Effekt	Indirekter Effekt	Korrelation
Verrichtungs-Gliederung	Baum vs. Tabelle	0.021	-0.09	-0.09	----	-0.02
	Gießen vs. Kiel 82		0.06	0.06	----	-0.01
	Gießen vs. Kiel 84		0.19	0.19	----	0.11
Objekt-Gliederung	Baum vs. Tabelle	0.121	0.31	0.31	----	0.11
	Gießen vs. Kiel 82		-0.06	-0.06	----	0.06
	Gießen vs. Kiel 84		-0.41	-0.41	----	-0.22
Vergleichstechnik	Baum vs. Tabelle	0.339	-0.03	-0.03	-0.01	0.21
	Gießen vs. Kiel 82		0.30	0.31	0.01	-0.00
	Gießen vs. Kiel 84		0.61	0.63	0.02	0.46
	Verrichtungs-Gliederung		0.25	0.25	----	0.29
	Objekt-Gliederung		0.06	0.06	----	-0.14
Aggregationsgrad	Baum vs. Tabelle	0.159	0.08	0.08	-0.00	-0.06
	Gießen vs. Kiel 82		-0.29	-0.29	0.00	-0.10
	Gießen vs. Kiel 84		-0.43	-0.42	0.01	-0.23
	Verrichtungs-Gliederung		0.21	0.21	-0.00	0.14
	Objekt-Gliederung		0.06	0.06	-0.00	0.08
	Vergleichstechnik		-0.01	-0.01	----	-0.14
Beschaffung	Baum vs. Tabelle	0.169	0.18	0.18	-0.01	0.30
	Gießen vs. Kiel 82		-0.13	-0.07	0.06	-0.22
	Gießen vs. Kiel 84		0.06	0.19	0.13	0.32
	Verrichtungs-Gliederung		-0.01	0.05	0.07	0.07
	Objekt-Gliederung		-0.01	0.01	0.02	-0.04
	Vergleichstechnik		0.23	0.23	-0.00	0.29
	Aggregationsgrad		0.04	0.04	----	-0.01
Übertragung	Baum vs. Tabelle	0.294	0.04	0.01	-0.03	0.05
	Gießen vs. Kiel 82		0.17	0.21	0.07	0.12
	Gießen vs. Kiel 84		-0.12	0.19	0.31	0.08
	Verrichtungs-Gliederung		0.03	0.16	0.13	0.23
	Objekt-Gliederung		-0.22	-0.19	0.03	-0.24
	Vergleichstechnik		0.36	0.41	0.05	0.40
	Aggregationsgrad		0.15	0.16	0.01	0.10
	Beschaffung		0.23	0.23	----	0.29
Kennzahlenbildung	Baum vs. Tabelle	0.263	-0.09	-0.02	0.07	0.06
	Gießen vs. Kiel 82		0.00	0.09	0.09	-0.00
	Gießen vs. Kiel 84		0.14	0.20	0.07	0.15
	Verrichtungs-Gliederung		-0.36	-0.30	0.05	-0.29
	Objekt-Gliederung		0.10	0.04	-0.06	0.08
	Vergleichstechnik		0.08	0.23	0.16	0.17
	Aggregationsgrad		-0.12	-0.06	0.06	-0.16
	Beschaffung		0.09	0.16	0.08	0.19
	Übertragung		0.33	0.33	----	0.27
Treffsicherheit der Gesamtlage	Baum vs. Tabelle	0.261	0.16	0.20	0.04	0.25
	Gießen vs. Kiel 82		0.22	0.04	-0.17	-0.08
	Gießen vs. Kiel 84		0.27	0.13	-0.14	0.21
	Verrichtungs-Gliederung		0.02	-0.10	-0.12	-0.08
	Objekt-Gliederung		-0.06	0.00	0.07	0.03
	Vergleichstechnik		-0.20	-0.20	-0.00	-0.05
	Aggregationsgrad		0.24	0.19	-0.05	0.10
	Beschaffung		0.23	0.21	-0.02	0.24
	Übertragung		-0.28	-0.19	0.09	-0.12
	Kennzahlenbildung		0.28	0.28	----	0.22

fluß der Übertragungsaktivität wird gemildert durch ihre positive Wirkung auf die effizienzsteigernde Kennzahlenbildung. Multipliziert man die beiden Pfadkoeffizienten, so erhält man gerade den indirekten Effekt der Übertragungsaktivität: $0,33 * 0,28 = 0,0924$. Bei der Informationsbeschaffung ergibt sich kein nennenswerter indirekter Effekt, so daß der direkte dem totalen Effekt entspricht.

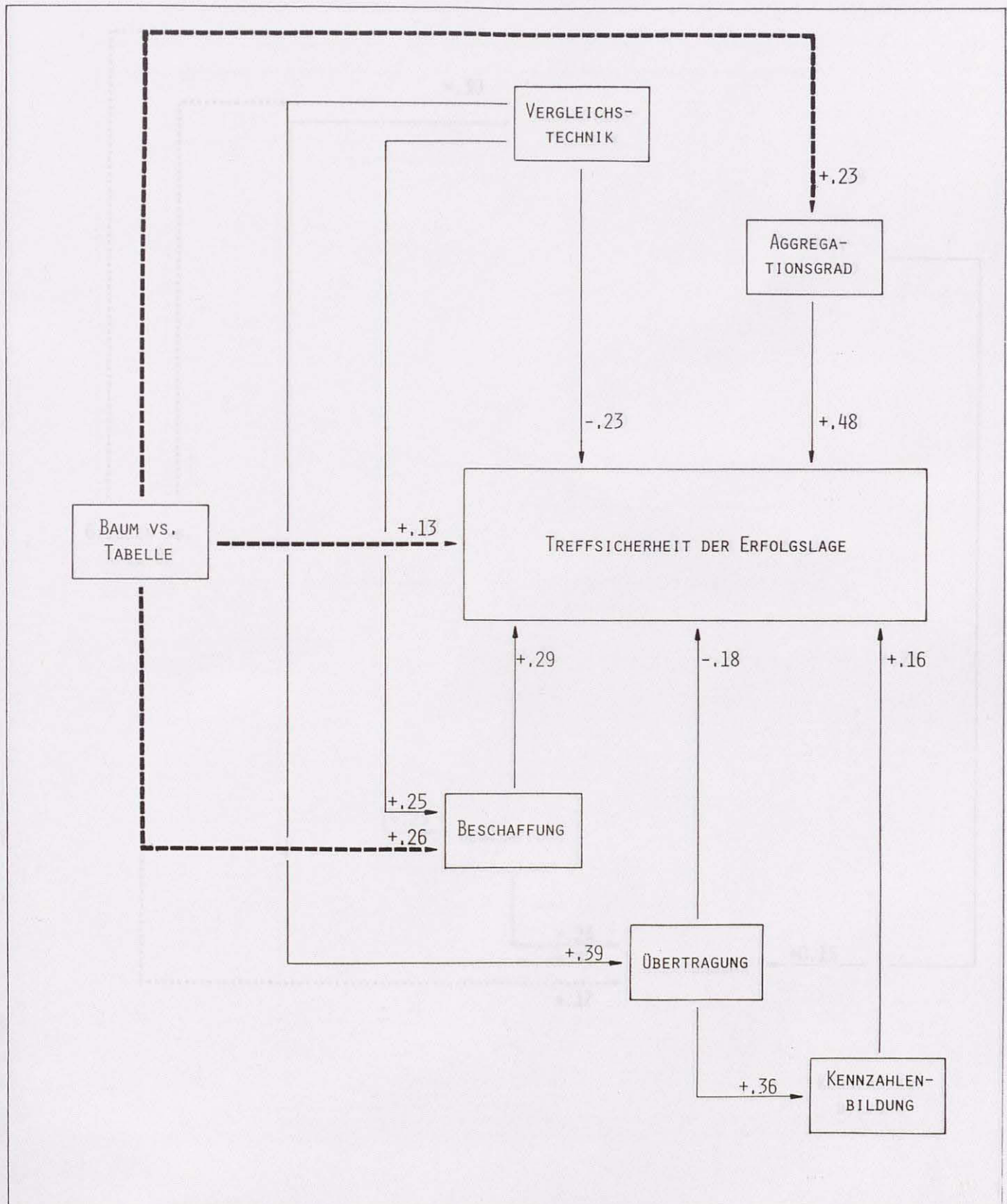
2. Taktische Dimensionen: Vergleichstechnik und Aggregationsgrad weisen (in der Summe) nur geringe indirekte Effekte auf. Die Vergleichstechnik übt zwar einen Einfluß von 0,36 auf die effizienzmindernde Übertragung aus, doch dieser Pfad wird kompensiert durch den positiven direkten Einfluß von 0,23 auf die effizienzsteigernde Beschaffungsaktivität. Anzumerken ist, daß sich beim Aggregationsgrad ein totaler Effekt von -0,20 für den quadratischen Term ergibt, wenn man die Pfadanalyse mit quadratischen Termen rechnet. Es wäre also falsch, aus dem hier ausgewiesenen linearen Modell zu schließen, daß ein möglichst hoher Anteil aggregierter Daten angestrebt werden sollte. Gute Analytiker zeichnen sich vielmehr durch einen **mittleren Anteil** aggregierter Daten aus.

Wie wirken die **Rahmenvariablen** auf dieses Netzwerk ein? Die Abbildungen 4.22. bis 4.24. illustrieren es anhand der wichtigsten Pfade:¹⁶⁾

3. Präsentationsform: Der totale Effekt der Baumstruktur ist nur etwas größer als der direkte Effekt, weil es lediglich einen kleinen indirekten Effekt gibt. Er wird hervorgerufen durch die vitalisierende Wirkung der Baumstruktur auf die Informationsbeschaffung.

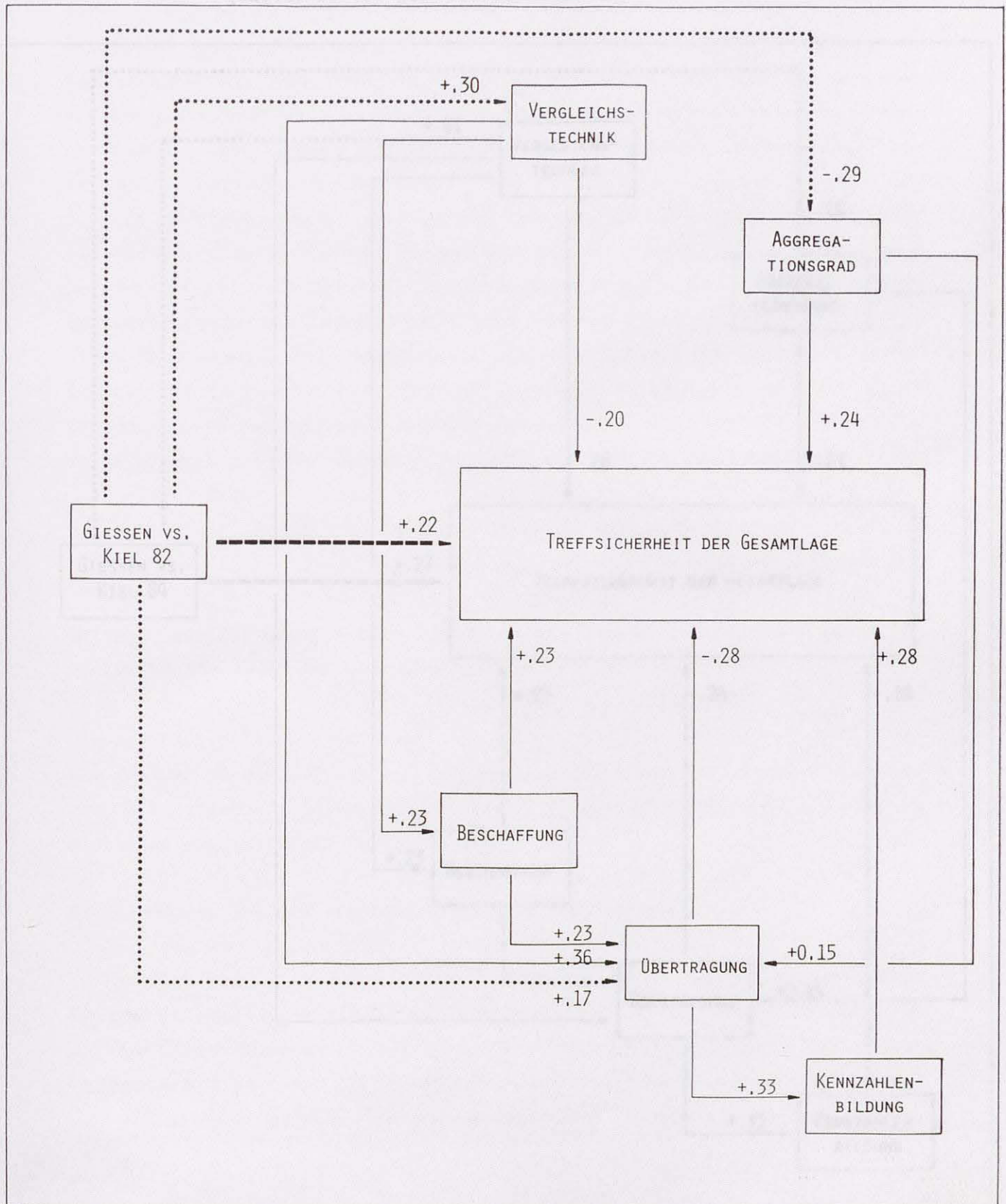
4. Vorbildung: Der Gesamteinfluß der Vorbildung ist **überraschend gering**: Der 0,22 betragende direkte Effekt des Kontrastes "Gießen vs. Kiel 82" wird durch einen beachtlichen indirekten Effekt auf einen vernachlässigbaren totalen Effekt von 0,04 reduziert. Der negative indirekte Effekt kann darauf zurückgeführt werden, daß die Gießener Analytiker eine komplexere Vergleichstechnik anwenden, einen geringeren Anteil aggregierter Daten nutzen und mehr Informationen übertragen (siehe Abbildung 4.23.).

Abbildung 4.22.: Interpersonelle Analyse – Einflüsse der Präsentationsform auf die Treffsicherheit der Gesamtlage (Basis: Je 103 gemittelte Meßwerte)



----- effizienzsteigernde Pfade der Präsentationsform

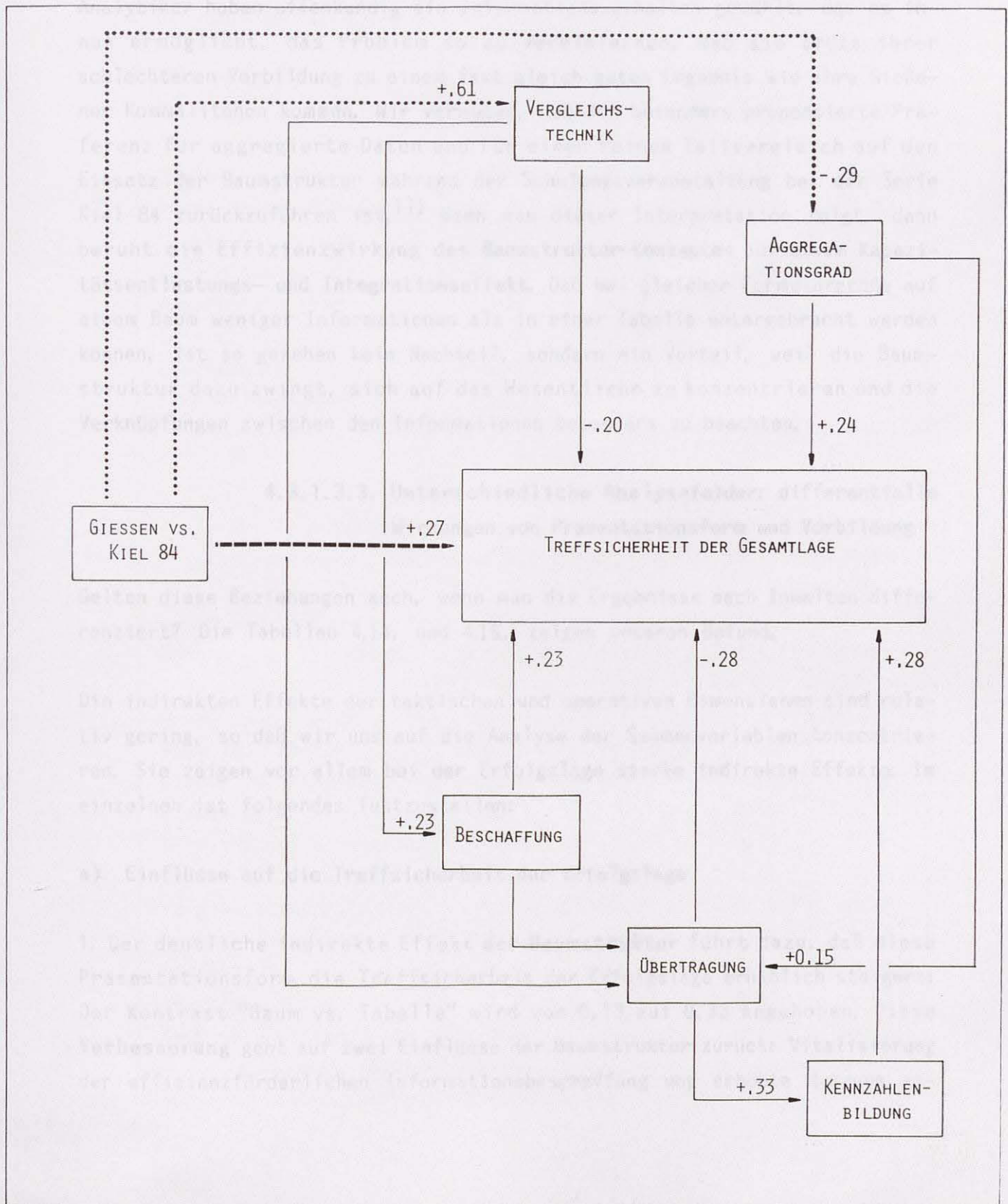
Abbildung 4.23.: Interpersonelle Analyse – Einflüsse der Vorbildung auf die Treffsicherheit der Gesamtlage (Basis: Je 103 gemittelte Meßwerte)



----- effizienzsteigernde Pfade der Vorbildung

..... effizienzsenkende Pfade der Vorbildung

Abbildung 4.24.: Interpersonelle Analyse – Einflüsse von Vorbildung und Präsentationsform auf die Treffsicherheit der Gesamtlage
(Basis: Je 103 gemittelte Meßwerte)



5. Der Kontrast "Gießen vs. Kiel 84" zeigt ebenfalls nur einen recht schwachen totalen Einfluß. Das Schrumpfen des Vorsprungs gegenüber den Kieler Analytikern beruht auf deren starker Präferenz für aggregierte Daten und deren einfacheren Vergleichstechnik (siehe Abbildung 4.24.). Die Kieler Analytiker haben offenkundig ein Informationsverhalten gewählt, das es ihnen ermöglicht, das Problem so zu vereinfachen, daß sie trotz ihrer schlechteren Vorbildung zu einem fast gleich guten Ergebnis wie ihre Gießener Kommilitonen kommen. Wir vermuten, daß die besonders prononcierte Präferenz für aggregierte Daten und für einen reinen Zeitvergleich auf den Einsatz der Baumstruktur während der Schulungsveranstaltung bei der Serie Kiel 84 zurückzuführen ist.¹⁷⁾ Wenn man dieser Interpretation folgt, dann beruht die **Effizienzwirkung des Baumstruktur-Konzeptes** auf einem **Kapazitätsentlastungs- und Integrationseffekt**. Daß bei gleicher Formulargröße auf einem Baum weniger Informationen als in einer Tabelle untergebracht werden können, ist so gesehen kein Nachteil, sondern ein Vorteil, weil die Baumstruktur dazu zwingt, sich auf das Wesentliche zu konzentrieren und die **Verknüpfungen** zwischen den Informationen besonders zu beachten.

4.3.1.3.3. Unterschiedliche Analysefelder: differentielle Wirkungen von Präsentationsform und Vorbildung

Gelten diese Beziehungen auch, wenn man die Ergebnisse nach Inhalten differenziert? Die Tabellen 4.14. und 4.15. zeigen unseren Befund.

Die indirekten Effekte der taktischen und operativen Dimensionen sind relativ gering, so daß wir uns auf die Analyse der Rahmenvariablen konzentrieren. Sie zeigen vor allem bei der Erfolgslage starke indirekte Effekte. Im einzelnen ist folgendes festzustellen:

a) Einflüsse auf die Treffsicherheit der Erfolgslage

1. Der deutliche indirekte Effekt der **Baumstruktur** führt dazu, daß diese Präsentationsform die Treffsicherheit der Erfolgslage erheblich steigert: Der Kontrast "Baum vs. Tabelle" wird von 0,13 auf 0,33 angehoben. Diese **Verbesserung** geht auf zwei Einflüsse der Baumstruktur zurück: Vitalisierung der effizienzförderlichen Informationsbeschaffung und erhöhte Nutzung ag-

Tabelle 4.14.: Interpersonelle Analyse – Pfadanalyse zur Treffsicherheit der Erfolgslage (Basis: Je 103 gemittelte Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korrelation
Baum vs. Tabelle	0.13	0.33	0.20	0.28
Gießen vs. Kiel 82	0.15	-0.12	-0.27	0.13
Gießen vs. Kiel 84	0.28	-0.16	-0.44	-0.07
Verrichtungs-Gliederung	0.01	-0.06	-0.07	-0.08
Objekt-Gliederung	0.01	0.14	0.13	0.19
Vergleichstechnik	-0.23	-0.18	0.05	-0.13
Aggregationsgrad	0.48	0.44	-0.05	0.39
Beschaffung	0.29	0.29	0.00	0.30
Übertragung	-0.18	-0.12	0.06	-0.10
Kennzahlenbildung	0.16	0.16	-----	0.15

Tabelle 4.15.: Interpersonelle Analyse – Pfadanalyse zur Treffsicherheit der Vermögens- und Finanzlage (Basis: Je 103 gemittelte Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korrelation
Baum vs. Tabelle	0.00	-0.06	-0.06	0.11
Gießen vs. Kiel 82	0.16	0.21	0.06	0.01
Gießen vs. Kiel 84	0.25	0.44	0.19	0.31
Verrichtungs-Gliederung	0.03	-0.11	-0.13	0.04
Objekt-Gliederung	-0.09	-0.14	-0.06	-0.18
Vergleichstechnik	-0.01	0.09	0.10	0.25
Aggregationsgrad	0.09	0.11	0.02	0.11
Beschaffung	0.14	0.24	0.10	0.34
Übertragung	-0.20	-0.12	0.09	0.18
Kennzahlenbildung	0.48	0.48	-----	0.54

gregierter Daten (vgl. Abbildung 4.25.). Da sowohl die Informationsbeschaffung als auch der Aggregationsgrad eine signifikant umgekehrt u-förmige Beziehung zur Treffsicherheit der Erfolgslage aufweist, wird der Effekt der Präsentationsform auch für das Modell mit den quadratischen Termen bestimmt. Es ergeben sich in der Summe die gleichen direkten, indirekten und totalen Effekte. Dabei werden die positiven Pfade über die linearen Terme leicht verstärkt. Über die quadratischen Terme laufen sehr schwache negativ wirkende Pfade. **Die Baumstruktur setzt demnach offenkundig am ansteigenden Ast der umgekehrt u-förmigen Funktion an.**

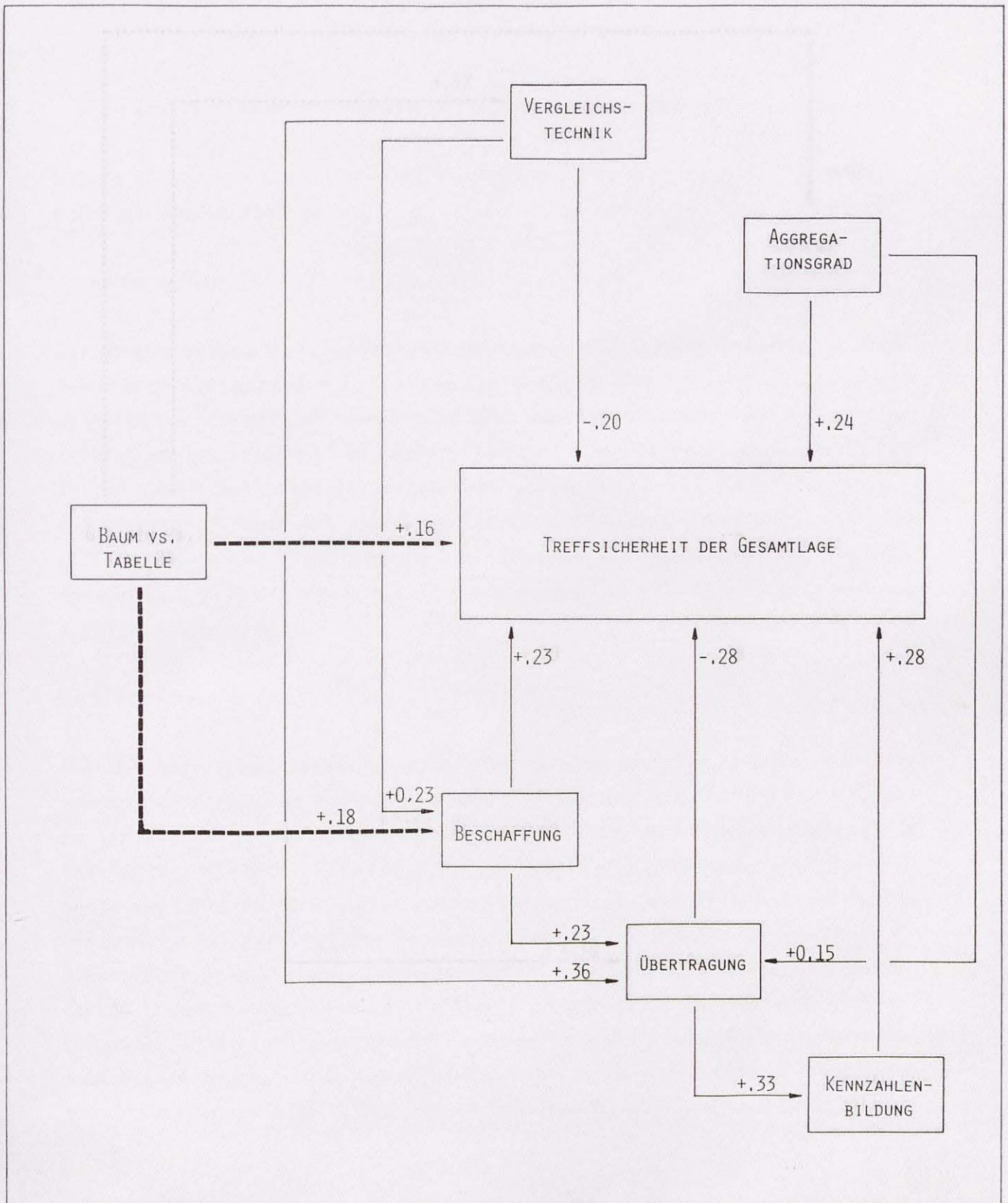
2. Die **Vorbildung** zeigt einen **negativen indirekten Effekt**: Der Kontrast "Gießen vs. Kiel 82" wird von +0,15 auf -0,12 verändert. Er beruht darauf, daß die Gießener Analytiker eine komplexere Vergleichstechnik anwenden, einen geringeren Anteil aggregierter Daten verarbeiten und weniger Informationen zur Erfolgslage beschaffen als die Kieler Probanden.

3. Dieser indirekte Effekt der unterschiedlichen Vorbildung ist bei der Wertung des indirekten Effekts des Kontrastes "Kiel vs. Gießen 84" zu berücksichtigen. Zieht man von dessen 0,44 die 0,27 des Vorbildungseffekts ab, so verbleiben 0,17, was in etwa dem Wert von 0,20 des Kontrastes "Baum vs. Tabelle" entspricht. Der indirekte Effekt beruht darauf, daß die Teilnehmer der Serie Kiel 84 einen wesentlich höheren Anteil aggregierter Informationen verarbeiten und konsequent darauf verzichten, Informationen über den Vergleichsfall zu analysieren, während die Gießener Analytiker solche Daten zwar beschaffen und auf ihre Analysebögen übertragen, aber bei der Bildung ihrer Urteile nur mangelhaft nutzen (vgl. Abbildung 4.26.).

b) Einflüsse auf die Treffsicherheit der Vermögens- und Finanzlage:

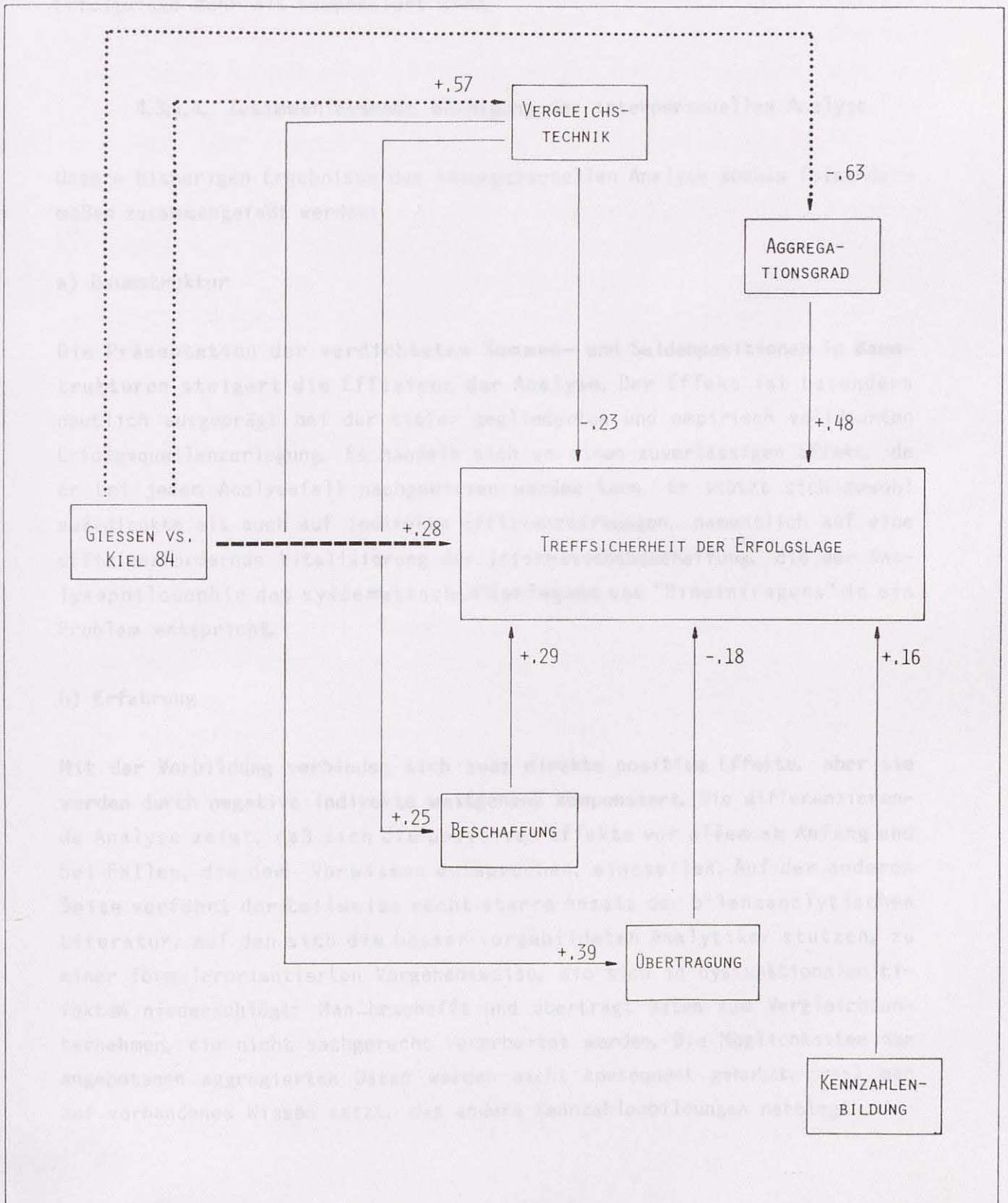
Bei diesem Analysefeld ergeben sich keine nennenswerten indirekten Effekte. Es zeigt sich lediglich, daß die Analytiker der Serie Kiel 84 die Vermögens- und Finanzlage gegenüber ihren Gießener Kollegen deutlich vernachlässigen. Dies dürfte ein Einfluß der geringeren Vorbildung sein, da sich beim Kontrast "Baum vs. Tabelle" kein solcher Effekt zeigt. Wie unsere Auswertung der bilanzanalytischen Monographien gezeigt hat, behandeln diese die

Abbildung 4.25.: Interpersonelle Analyse – Einflüsse der Präsentationsform auf die Treffsicherheit der Erfolgslage (Basis: Je 103 gemittelte Meßwerte)



----- effizienzsteigernde Pfade der Präsentationsform

Abbildung 4.26.: Interpersonelle Analyse – Einflüsse von Vorbildung und Präsentationsform auf die Treffsicherheit der Erfolgslage
(Basis: Je 103 gemittelte Meßwerte)



----- effizienzsteigernde Pfade des Kontrastes "Gießen vs. Kiel 84"

..... effizienzsenkende Pfade des Kontrastes "Gießen vs. Kiel 84"

Vermögens- und Finanzlage intensiver als die Erfolgslage. Für die Treffsicherheit der Gesamtlage ist festzustellen, daß die Vernachlässigung der Vermögens- und Finanzlage durch die wesentlich intensivere Bearbeitung der Erfolgslage mehr als kompensiert wird.

4.3.1.4. Zusammenfassende Würdigung der interpersonellen Analyse

Unsere bisherigen Ergebnisse der interpersonellen Analyse können folgendermaßen zusammengefaßt werden:

a) Baumstruktur

Die Präsentation der verdichteten Summen- und Saldenpositionen in Baumstrukturen steigert die Effizienz der Analyse. Der Effekt ist besonders deutlich ausgeprägt bei der tiefer gegliederten und empirisch validierten Erfolgsquellenzerlegung. Es handelt sich um einen zuverlässigen Effekt, da er bei jedem Analysefall nachgewiesen werden kann. Er stützt sich sowohl auf direkte als auch auf indirekte Effizienzwirkungen, namentlich auf eine effizienzfördernde Vitalisierung der Informationsbeschaffung, die der Analysephilosophie des systematischen Zerlegens und "Hineinfragens" in ein Problem entspricht.

b) Erfahrung

Mit der Vorbildung verbinden sich zwar direkte positive Effekte, aber sie werden durch negative indirekte weitgehend kompensiert. Die differenzierende Analyse zeigt, daß sich die positiven Effekte vor allem am Anfang und bei Fällen, die dem Vorwissen entsprechen, einstellen. Auf der anderen Seite verführt der teilweise recht starre Ansatz der bilanzanalytischen Literatur, auf den sich die besser vorgebildeten Analytiker stützen, zu einer formularorientierten Vorgehensweise, die sich in dysfunktionalen Effekten niederschlägt: Man beschafft und überträgt Daten zum Vergleichsunternehmen, die nicht sachgerecht verarbeitet werden. Die Möglichkeiten der angebotenen aggregierten Daten werden nicht konsequent genutzt, weil man auf vorhandenes Wissen setzt, das andere Kennzahlenbildungen nahelegt.

c) Informationsverhalten

Die für alle Dimensionen des Informationsverhaltens postulierte umgekehrt u-förmige Beziehung kann in dieser generellen Form nicht aufrecht erhalten werden. Unsere Befunde zeigen lediglich bei der Informationsbeschaffung für ungeübte Analytiker und beim Aggregationsgrad für bestimmte Analysefälle einen signifikant umgekehrt u-förmigen Verlauf. Bei den übrigen fünf Dimensionen zeigt auch eine Differenzierung nach bearbeiteten Fällen und analysierten Feldern kein hypothesenkonformes Ergebnis. Die Ursache sehen wir in der großen Heterogenität der Analytiker: Es gibt offenkundig unterschiedliche Möglichkeiten, das Informationsproblem Bilanzanalyse zufriedenstellend zu lösen. Eine für alle Analytiker "optimale" Verhaltensweise läßt sich aus unseren Befunden nicht ableiten. Vor vereinfachenden Rezepten muß gewarnt werden.

Auf der anderen Seite sollte man den **diagnostischen Wert** der ermittelten Beziehungen nicht unterschätzen. Die erklärten Varianzen betragen 38,4 bzw. 37,4 v. H. für die Treffsicherheit der Erfolgslage bzw. der Vermögens- und Finanzlage, und zwar bei den robusten linearen Modellen. Man kann daher durchaus Empfehlungen ableiten. Der Blick sollte sich jedoch nicht auf das Erreichen einer optimalen Lösung richten. Es gilt vielmehr **Insuffizienzen** zu vermeiden. Hierfür bieten sich folgende Anknüpfungspunkte an:

1. Positiv wirkende Aktivitäten können gefördert, negativ wirkende begrenzt werden. So sollte man die Analytiker zur Nutzung verdichteter Informationen und zur Bildung von Kennzahlen anregen und vor einer mechanistischen Anwendung komplexer Vergleichstechniken warnen. Wichtig ist dabei nicht die Tatsache, daß Kennzahlen berechnet werden, sondern daß man **sachgerechte Relativierungen** vornimmt und **Beziehungen** zwischen verschiedenen Positionen herstellt, um typische Muster zu identifizieren.

2. **Extreme Ausprägungen** einzelner Verhaltensdimensionen setzen **Schranken** für die erreichbare Effizienz, wie das Streupunktdiagramm zur Verrichtungsgliederung exemplarisch belegt: Bei einer extrem geringen Verrichtungsgliederung werden sehr kleine Informationspäckchen beschafft und unmittelbar

verarbeitet, so daß sich im Ergebnis eine beziehungslose Aneinanderreihung stereotyper Einzelaussagen ergibt. Eine extrem hohe Verrichtungsgliederung führt dazu, daß viele Daten beschafft und übertragen werden, die fallspezifisch irrelevant sind oder aus Zeitmangel nicht mehr verarbeitet werden. Unsere Hypothese: "Wenn eine extreme Verrichtungsgliederung gewählt wird, stellt sich eine geringe Effizienz ein", wird durch den Befund **nicht** widerlegt. **Unzulässig** ist nur der **Umkehrschluß**: Wer eine mittlere Verrichtungsgliederung wählt, kann aus **anderen** Gründen trotzdem eine schlechte Analyse anfertigen.

Der Wert unseres Diagnoseinstruments liegt darin, daß sich wenige Analytiker bei **allen** Dimensionen zweckmäßig verhalten. Man wird daher vielen Analytikern helfen können, indem man sie gezielt auf ihre auf **individuelle** Insuffizienzen hinweist. Ob man für diese Insuffizienzanalyse auch ein baumstrukturartiges Kennzahlensystem entwickelt, ist ein weiterführender Gedanke, dessen Konsequenzen hier nicht diskutiert werden sollen.

3. Wir vermuten, daß bei schlechten Analytikern mehrere Insuffizienzen zusammentreffen und ein **Syndrom** bilden. Häufig auftretende Syndrome können mit Hilfe einer Clusteranalyse unserer Verhaltensdimensionen identifiziert werden. Die so gebildeten Typen können zu Persönlichkeitsmerkmalen in Beziehung gesetzt werden, um Ursachen der Insuffizienzen zu erforschen. Auswirkungen zur Effizienz können dysfunktionale Wechselwirkungen der Verhaltensdimensionen erkunden. Da unsere Untersuchung hypothesengeleitet ist, haben wir diesen Schritt noch nicht vollzogen.

4. Die Auswertungen zum Informationsniveau stützen sich auf 112 Informationsprozesse statt wie bisher auf 103 Versuchspersonen. Wegen der höheren Teststärke werden schwächere Effekte statistisch signifikant. Wir konnten trotzdem das Signifikanzniveau bei $\alpha = 0,05$ gleiches Akzeptanzniveau für die Abschätzung gegenüber möglichen Stichprobeneffekten beibehalten.

4.3.2. Die intertemporale Analyse: das Lernverhalten der Analytiker

4.3.2.1. Zur Methodik der intertemporalen Analyse

Das intertemporale Testdesign weist einige Besonderheiten auf, die kurz erläutert werden sollen, um die Befundpräsentation zu entlasten.

1. Bei der intertemporalen Analyse werden ipsative Meßwerte verwendet. Sie erfassen drei verschiedene Einflüsse: Die Einwirkungen unterschiedlicher Analysefälle, die Effekte von Einübung und Selbststeuerung sowie die Wirkungen der bilanzanalytischen Feedbackveranstaltungen. Die starken Veränderungen des Informationsverhaltens und die stetige Zunahme der Effizienz der Analyse veranlassen uns, diese verschiedenen Einflüsse unter dem Begriff **Lernverhalten** zu subsumieren.

2. Bei der interpersonellen Analyse haben wir Analytiker mit unterschiedlicher Motivation und Fähigkeit verglichen. Nun wird das Verhalten bei Anwendung eines spezifischen **Lehrprogramms** und einer besonderen, aus didaktischen Gründen gewählten **Fallsequenz** betrachtet. Unsere Aussagen können nur für dieses Lehrprogramm Geltung beanspruchen. Wir untersuchen, **wie** und **was** gelernt wird, und nicht, ob diese Veranstaltung möglichen anderen Trainingsprogrammen unter- oder überlegen ist. Daß das gewählte Programm wirksam ist, wird im einzelnen belegt.

3. Im Mittelpunkt der intertemporalen Analyse stehen die Beziehungen zwischen der Entwicklung des Informationsverhaltens und der Entwicklung der Effizienz der Analyse. Ob sich bei unterschiedlichen **Präsentationsformen** oder unterschiedlicher **Vorbildung** andere Beziehungen ergeben, ist eine Frage, die mit Hilfe von **Interaktionstests** zu überprüfen ist. Sie werden erst im Anschluß an die Hauptanalyse vorgenommen.¹⁸⁾

4. Die Auswertungen zum Lernverhalten stützen sich auf 412 Informationsprozesse statt wie bisher auf 103 Versuchspersonen. Wegen der höheren **Teststärke** werden schwächere Einflüsse statistisch signifikant. Wir behalten trotzdem das Signifikanzniveau bei, um ein gleichartiges **Akzeptanzkriterium** für die Absicherung gegenüber möglichen Stichprobeneffekten anzuwen-

den. Um einer Interpretation praktisch unbedeutender Effekte vorzubeugen, tritt das hiermit vorab (!) festgelegte **Relevanzkriterium** hinzu, daß die Regressionskoeffizienten in unserer Stichprobe wenigstens 0,10 betragen sollen.

5. Bortz weist darauf hin, daß das Rechnen mit ipsativen Meßwerten einer Varianzanalyse mit Meßwiederholungen entspricht, wenn man bei der Berechnung und Interpretation der Teststatistiken die **Zahl der Freiheitsgrade** um die Zahl der Versuchspersonen verringert.¹⁹⁾ Wir folgen seiner Empfehlung, weil wir die Information, die in den interindividuellen, über die Fälle gemittelten Unterschieden enthalten ist, bereits bei der interpersonellen Analyse genutzt haben.

6. Schließlich ist auf ein wichtiges **Interpretationsproblem** hinzuweisen: Die Veränderungen des Informationsverhaltens und die Veränderungen der Effizienz beruhen zu einem nicht unbeachtlichen Teil auf einer **gemeinsamen Ursache**: dem bilanzanalytischen Feedback, das während der Schulungsveranstaltungen vermittelt wird. Es ist nicht unbedenklich, eine zeitliche Kovariation von Informationsverhalten und Effizienz kausal zu interpretieren. Das Problem der kausalen Interpretation stellt sich natürlich auch bei der interpersonellen Analyse, aber dort gibt es weitaus weniger Gründe anzunehmen, die interindividuellen Unterschiede in ähnlicher Weise auf ein einziges Persönlichkeitsmerkmal zurückzuführen. Die interpersonellen Abweichungen, die durch den experimentell variierten Faktor Präsentationsform zu erklären sind, können kausal interpretiert werden, da die interne Validität des Experiments gesichert ist. Dies gilt nun nicht mehr. Eine kausale Interpretation wird erst möglich sein, wenn man systematisch unterschiedliche Trainingsprogramme variiert. Dieser zweite Schritt kann jedoch auf unserem ersten aufbauen und unsere Aussagen als **Hypothesen** begreifen.

4.3.2.2. Der Ausgangsbefund: Erneute Falsifikation der umgekehrt u-förmigen Verlaufshypothese

Kann unsere umgekehrt u-förmige Verlaufshypothese zur Effizienz des Informationsverhaltens in dem Sinn Geltung beanspruchen, daß bei jedem einzelnen Analytiker eine Abweichung von seinem persönlichen mittleren Verhalten zu einer Effizienzminderung führt?

Tabelle 4.16. beantwortet diese Frage mit einem klaren Nein: Keiner der quadratischen Terme der Dimensionen des Informationsverhaltens erreicht die geforderte Irrtumswahrscheinlichkeit. Die Regressionskoeffizienten liegen dicht bei Null. Die umgekehrt u-förmige Verlaufshypothese ist auch im intertemporalen Test als widerlegt anzusehen.

Die Datenanalyse zeigt, daß die Effizienz signifikant wächst mit:

- * verstärkter Informationsbeschaffung,
- * intensiverer Nutzung aggregierter Daten,
- * vermehrter Kennzahlenbildung und
- * vereinfachter Vergleichstechnik.

Dieses Ergebnis deckt sich mit unseren Beobachtungen bei der interpersonellen Analyse. Auch diesmal zeigt sich kein nennenswerter direkter Effekt der Gliederung des Ablaufs des Informationsprozesses nach Objekten oder Einrichtungen. Die Effizienz der Analyse nimmt mit einer Einschränkung der Übertragungsaktivität leicht ab, aber dieser Effekt ist schwach und nicht signifikant.

Trotz der Parallelität der Befunde ist zu prüfen, ob die unabhängigen Variablen weit genug streuen, d. h. ob das Informationsverhalten bezüglich seiner Veränderung im Zeitablauf eine hinreichende Variation aufweist.

Tabelle 4.16.: Intertemporale Analyse – Treffsicherheit der Gesamtlage
(Basis: Je 412 ipsative Meßwerte)

Explikative Variablen	Hypo- these	Regressionskoeffizienten und Signifikanzniveaus			
		Linearer Term		Quadr. Term	
		Beta	Sign.&	Beta	Sign.&
Strategische Dimensionen					
Objekt-Gliederung	n	0,09	0,15	-0,08	0,15
Verrichtungs-Gliederung	n	-0,07	0,23	-0,02	0,74
Taktische Dimensionen					
Vergleichstechnik	n	-0,13	0,02	-0,03	0,58
Aggregationsgrad	n	0,18	0,00	-0,01	0,86
Operative Dimensionen					
Beschaffung	n	0,14	0,01	-0,01	0,85
Übertragung	n	-0,09	0,11	0,00	0,94
Kennzahlenbildung	n	0,18	0,00	-0,02	0,73

Erklärte Varianz: 22,1 % Signifikanz: $p = 0,00$

&: Zweiseitiger Test, wegen der ipsativen Meßwerte t-Statistik und Freiheitsgrade um die Anzahl der Versuchspersonen vermindert

n: Umgekehrt u-förmiger Zusammenhang

4.3.2.3. Die Entwicklung des Informationsverhaltens: unterschiedlich ausgeprägt bei den einzelnen Dimensionen

Wie entwickelt sich das Informationsverhalten während des Lernprozesses? Welche Dimensionen verändern sich, welche bleiben konstant? Wie gleichförmig oder wie unterschiedlich vollzieht sich dieser Wandlungsprozeß bei den einzelnen Analytikern?

Tabelle 4.17. zeigt die mittleren Ausprägungen der ipsativen Meßwerte bei den vier Analysefällen (Datenzeile "MW"). Sie gibt an, wie stark die ipsativen Meßwerte der Analytiker um diese mittlere Abweichung vom individuellen Mittelwert streuen (Datenzeile "STD"). Schließlich wird dargestellt, um wieviel Prozent die ipsativen Meßwerte vom mittleren Wert aller Informationsprozesse und Analytiker abweichen (Datenzeile "ABW"). Mit dieser Relativierung sollen die unterschiedlichen Dimensionen des Informationsverhaltens vergleichbar gemacht werden.

Die Darstellung wird am Beispiel der Informationsbeschaffung erläutert:

- * Der Wert von -15,1 in der Datenzeile "MW" besagt, daß für die Analyse des ersten Falls durchschnittlich 15,1 weniger Informationen als im Mittel aller Informationsprozesse beschafft werden.
- * Um diese mittlere Abweichung streuen die Werte der einzelnen Analytiker mit einer Standardabweichung von 24,9.
- * Die in der dritten Datenzeile angegebene mittlere prozentuale Abweichung errechnet sich als 100 mal mittlere absolute Abweichung dividiert durch den Mittelwert aller Informationsprozesse, das heißt im gewählten Beispiel - $15,1 * 100 : 90,4 = -16,7$.²⁰⁾

Das Beispiel illustriert, wie unterschiedlich sich das Informationsverhalten bei den Analytikern entwickelt. Daß auch bei den anderen Dimensionen eine erhebliche Variation besteht, kann man daran ablesen, daß die Standardabweichungen der ipsativen Meßwerte meist deutlich größer als ihre

Tabelle 4.17.: Intertemporale Analyse - Entwicklung des Informationsverhaltens (Basis: 412 ipsative Meßwerte)

Explikative Variablen		Meßwerte bei den Analysefällen			
		1	2	3	4
Strategische Dimensionen					
Objekt-Gliederung ^{&}	MW	-26,0	6,2	9,7	10,2
	STD	23,7	18,0	14,7	13,6
	ABW	-39,2	9,8	14,6	15,4
Verrichtungs-Gliederung ^{&}	MW	3,1	1,8	-2,6	-2,3
	STD	10,5	9,4	8,3	8,2
	ABW	4,7	2,7	-3,7	-3,6
Taktische Dimensionen					
Vergleichstechnik	MW	3,9	1,1	-1,5	-3,5
	STD	8,4	6,4	5,3	5,6
	ABW	24,0	6,8	-9,3	-21,6
Aggregationsgrad	MW	-14,4	2,0	6,2	6,3
	STD	14,4	11,9	10,1	12,3
	ABW	-24,0	3,3	10,4	20,5
Operative Dimensionen					
Beschaffung	MW	-15,1	1,4	3,9	9,8
	STD	24,9	13,5	16,1	15,3
	ABW	-16,7	1,5	4,3	10,8
Übertragung	MW	0,9	1,1	-1,2	-0,8
	STD	22,2	16,8	15,3	21,2
	ABW	2,3	2,9	-4,3	-2,1
Kennzahlenbildung	MW	-2,8	-1,3	2,4	1,8
	STD	8,5	6,6	6,8	7,2
	ABW	-14,1	-6,6	12,1	9,1

MW: Mittelwert STD: Standardabweichung ABW: Prozentuale Abweichung

&: Ursprüngliche Meßwerte zur besseren Vergleichbarkeit der Dimensionen auf das Intervall von 0 bis 100 transformiert

Mittelwerte sind. Es besteht daher kein hinreichender Grund, die Falsifikation der umgekehrt u-förmigen Verlaufshypothesen anzuzweifeln.

Bei inhaltlicher Würdigung des Befunds sind folgende Tendenzen hervorzuheben: Die Analytiker folgen dem Beispiel des Vortragenden und beginnen zunehmend mit der Erfolgslage. Sie vereinfachen ihre Vergleichstechnik und nutzen verstärkt aggregierte Daten. Mittelmäßig ausgeprägte Veränderungen ergeben sich bei der Informationsbeschaffung und Kennzahlenbildung. Diese Entwicklungen stehen in einem direkten Zusammenhang mit der Effizienz der Analyse. Ob man sie als "Ursache" für die Effizienzsteigerung ansieht, ist eine weiterführende Frage, deren Beantwortung zurückgestellt werden soll.

4.3.2.4. Differenzierung nach Analysefeldern: engerer Zusammenhang bei der Erfolgslage

Gelten die festgestellten Zusammenhänge für beide Analysefelder oder zeigen sich auch beim Lernverhalten deutliche Unterschiede zwischen der Erfolgslage und der Vermögens- und Finanzlage?

Tabelle 4.18. zeigt, daß bei der Erfolgslage ein engerer Zusammenhang besteht: Die erklärte Varianz beträgt bei ihr 43,1 v. H., bei der Vermögens- und Finanzlage sind es nur 17,1 v. H. Auch nach dieser Befunddifferenzierung sind unsere umgekehrt u-förmigen Hypothesen zur Effizienzwirkung des Informationsverhaltens als falsifiziert anzusehen. Lediglich bei der Informationsbeschaffung zeigt sich für die Vermögens- und Finanzlage ein signifikanter quadratischer Regressionskoeffizient, der gerade die gesetzte Relevanzschwelle von 0,10 erreicht (einseitiger hypothesenbezogener Test). Doch dieser Effekt wiegt zu gering, um die generelle Hypothese aufrechterhalten zu können. Sie wird damit definitiv als gescheitert erklärt.

Im einzelnen zeigen sich folgende signifikante und relevante Effekte:

1. **Eine Intensivierung der Kennzahlenbildung steigert die Effizienz.** Diese Aussage gilt sowohl für die Erfolgslage als auch für die Vermögens- und Finanzlage. Dort ist der Effekt besonders stark ausgeprägt.

Tabelle 4.18.: Intertemporale Analyse – Effizienzwirkungen differenziert nach Inhalten (Basis: Je 412 ipsative Meßwerte)

Explikative Variablen		Regressionskoeffizienten		
		Gesamt- lage	Erfolgs- lage	Vermögens- und Finanzlage
Strategische Dimensionen				
Objekt-Gliederung	Lin.	0,09	0,12 [*]	-0,02
	Quad.	-0,08	-0,06	-0,04
Verrichtungs-Gliederung	Lin.	-0,07	-0,08	-0,02
	Quad.	-0,02	-0,06	0,00
Taktische Dimensionen				
Vergleichstechnik	Lin.	-0,13 [*]	-0,17 ^{**}	0,06
	Quad.	-0,03	-0,01	0,02
Aggregationsgrad	Lin.	0,18 ^{**}	0,23 ^{**}	-0,01
	Quad.	-0,01	-0,03	-0,04
Operative Dimensionen				
Beschaffung	Lin.	0,14 ^{**}	0,29 ^{**}	0,00
	Quad.	-0,01	-0,04	-0,10 ⁺
Übertragung	Lin.	-0,09	-0,06	0,07
	Quad.	0,00	-0,06	0,08
Kennzahlenbildung	Lin.	0,18 ^{**}	0,17 ^{**}	0,36 ^{**}
	Quad.	-0,02	-0,01	-0,04
Erklärte Varianz:		22,1 % ^{**}	43,1 % ^{**}	17,1 % ^{**}

Signifikanzniveaus: + = 10 %, * = 5 %, ** = 1 % Niveau, zweiseitiger Test, wegen der ipsativen Meßwerte t-Statistik und Freiheitsgrade um die Zahl der Versuchspersonen verringert

2. Die **positive Wirkung der Informationsbeschaffung** zeigt sich nur bei der Erfolgslage, dort aber besonders stark.
3. Eine **intensivere Nutzung aggregierter Daten** führt gleichfalls nur bei der Erfolgslage zu einer **Effizienzsteigerung**. Auch in diesem Fall ist der Effekt vergleichsweise deutlich ausgeprägt.
4. Die negative Wirkung einer zu komplexen Vergleichstechnik oder besser: die **positiven Effekte der faktisch erfolgten Vereinfachung der Vergleichstechnik** gelten ebenfalls nur bei der Erfolgslage.
5. Schließlich zeigt sich ein signifikanter, aber schwächer ausgeprägter **positiver Einfluß der Veränderung der Objektgliederung**: Analytiker, die (in stärkerem Maße) dazu übergehen, die Erfolgslage vor der Vermögens- und Finanzlage zu untersuchen, steigern ihre Treffsicherheit bei der Beurteilung der Erfolgslage. Dieser Effekt mag nicht sonderlich überraschen, da eine frühere Bearbeitung der Erfolgslage in engem Zusammenhang mit einer intensiveren Beachtung dieses Analysefelds steht. Erstaunlich ist jedoch, daß sich bei der Vermögens- und Finanzlage **keine Senkung** ergibt. Wir werden zu prüfen haben, ob dieser erstmals festgestellte Einfluß der Ablaufgestaltung des Informationsprozesses durch indirekte Effekte verstärkt wird.

Insgesamt gesehen weist die Befunddifferenzierung nach Analysefeldern starke Parallelen zur interpersonellen Analyse auf. Die dort aufgestellten Erklärungen werden bestätigt. Überraschend ist allerdings, daß sich bei der Vermögens- und Finanzlage eine vergleichsweise geringe Varianzerklärung ergibt. Wir legen daher einen Ergänzungsbefund vor, der zeigt, wie sich das Informationsverhalten bei beiden Analysefeldern entwickelt (vgl. Tabelle 4.19.). Im Interesse einer übersichtlichen Darstellung werden diesmal nur die prozentualen Veränderungen der Meßwerte, relativiert am Mittelwert aller 412 Informationsprozesse, ausgewiesen.

Tabelle 4.19.: Intertemporale Analyse – Entwicklung der Dimensionen des Informationsverhaltens, differenziert nach Inhalten
(Basis: 412 ipsative Meßwerte)

Verhaltensdimensionen	Inhalte	Meßwerte bei den Analysefällen			
		1	2	3	4
Strategische Dimensionen Verrichtungs-Gliederung ^{&}	Erfolg	2,5	3,5	-3,3	-2,7
	VerFin	6,8	6,1	-9,6	-3,3
Taktische Dimensionen Vergleichstechnik	Erfolg	9,6	13,5	-8,3	-15,4
	VerFin	32,5	2,3	-10,3	-24,7
Aggregationsgrad	Erfolg	-34,8	4,4	16,0	14,4
	VerFin	-8,9	0,1	1,1	7,6
Operative Dimensionen Beschaffung	Erfolg	-29,3	4,2	7,8	16,9
	VerFin	-4,3	1,2	0,6	4,8
Übertragung	Erfolg	-24,8	10,4	7,1	7,5
	VerFin	32,0	-5,5	-14,3	-12,1
Kennzahlenbildung	Erfolg	-47,0	-0,9	26,0	21,6
	VerFin	25,3	-13,0	-5,4	-6,8

Erfolg: Erfolgslage, VerFin: Vermögens- und Finanzlage

&: Ursprüngliche Meßwerte zur besseren Vergleichbarkeit der Dimensionen auf das Intervall von 0 bis 100 transformiert

Bei der **Erfolgslage** verändern sich die Dimensionen Vergleichstechnik, Aggregationsgrad, Beschaffung und Kennzahlenbildung, die einen signifikant positiven direkten Effekt aufweisen, besonders stark. Die Verrichtungsgliederung, die keinen Bezug zur Leistungsentwicklung aufweist, bleibt praktisch konstant. Überraschend ist, daß bei der Übertragung kein nennenswerter direkter Einfluß auf die Effizienz der Analyse besteht, obwohl sich diese Aktivität deutlich verändert. Es gibt hier möglicherweise interindividuell unterschiedlich ausgeprägte Effizienzwirkungen, die sich gegenseitig kompensieren.

Bei der **Vermögens- und Finanzlage** wird die Vergleichstechnik stark vereinfacht und die Übertragungsaktivität deutlich gesenkt. Dies hat aber keinen direkten Einfluß auf die Effizienz der Analyse. Nur die spürbare Einschränkung der Kennzahlenbildung steht in einem direkten Zusammenhang mit der Treffsicherheit der Beurteilung der Vermögens- und Finanzlage. Der Rückgang dieser Effizienzgröße ist jedoch wesentlich schwächer als der gleichzeitige Anstieg der Treffsicherheit der Erfolgslage, so daß sich für die Treffsicherheit der Gesamtlage eine signifikante und stetige Verbesserung ergibt.

Da die wesentlichen Lernfortschritte bei der Beurteilung der Erfolgslage stattfinden, wollen wir die anschließende Pfadanalyse auf dieses Analysefeld konzentrieren. Aus den gleichen Gründen wie bei der interpersonellen Analyse beschränken wir uns auch diesmal auf ein lineares Modell.

4.3.2.5. Pfadanalyse der Effizienzwirkungen: Die Veränderung des Analysekonzeptes als Ursache des Lernfortschritts

4.3.2.5.1. Gemeinsame Analyse aller Informationsprozesse: beachtliche Effekte der Objektgliederung

Tabelle 4.20 zeigt einen beachtlichen Einfluß der Objektgliederung: Die positive direkte Effizienzwirkung wird von +0,12 durch einen indirekten Effekt von +0,28 auf einen totalen Effekt von +0,40 angehoben!

Tabelle 4.20.: Intertemporale Analyse – Pfadanalyse zur Treffsicherheit der Erfolgslage (Basis: Je 412 ipsative Meßwerte)

Abhängige Variablen	Unabhängige Variablen	Erklärte Varianz	Direkter Effekt	Totaler Effekt	Indirekter Effekt	Korrelation
Vergleichstechnik	Verrichtungsgliedg. Objektgliederung	0.025	0.16 0.01	0.16 0.01	---- ----	0.16 -0.02
Aggregationsgrad	Verrichtungsgliedg. Objektgliederung Vergleichstechnik	0.214	-0.05 0.45 0.04	-0.04 0.45 0.04	0.01 0.00 ----	-0.13 0.46 0.02
Beschaffung	Verrichtungsgliedg. Objektgliederung Vergleichstechnik Aggregationsgrad	0.201	-0.07 0.31 0.16 0.16	-0.05 0.38 0.16 0.16	0.02 0.07 0.01 ----	-0.12 0.39 0.15 0.31
Übertragung	Verrichtungsgliedg. Objektgliederung Vergleichstechnik Aggregationsgrad Beschaffung	0.254	0.26 0.20 0.21 0.11 0.20	0.28 0.33 0.25 0.14 0.20	0.02 0.13 0.04 0.03 ----	0.21 0.27 0.28 0.23 0.31
Kennzahlenbildung	Verrichtungsgliedg. Objektgliederung Vergleichstechnik Aggregationsgrad Beschaffung Übertragung	0.213	-0.11 0.25 -0.08 0.10 0.12 0.12	-0.10 0.38 -0.03 0.13 0.15 0.12	0.01 0.13 0.05 0.04 0.02 ----	-0.17 0.40 -0.05 0.29 0.29 0.20
Treffsicherheit der Erfolgslage	Verrichtungsgliedg. Objektgliederung Vergleichstechnik Aggregationsgrad Beschaffung Übertragung Kennzahlenbildung	0.414	-0.05 0.12 -0.18 0.25 0.29 -0.05 0.20	-0.14 0.40 -0.14 0.32 0.31 -0.03 0.20	-0.09 0.28 0.04 0.06 0.01 0.02 ----	-0.22 0.43 -0.17 0.45 0.44 0.11 0.41

Wie kommt dieser unerwartet starke indirekte Effekt zustande? Auf welchen Pfaden beruht er? Sehen wir uns hierzu Abbildung 4.27. an: Sie enthält alle Pfade, die einen Absolutbetrag des direkten Effektes von wenigstens 0,10 aufweisen. Die Pfade, die mit der Übertragung zusammenhängen, werden vernachlässigt, da Veränderungen dieser Aktivität die Entwicklung der Effizienz kaum beeinflussen.

Die positive Wirkung der Veränderung der Ablauforganisation gründet sich auf drei Einflüsse: Versuchspersonen, die dazu übergehen, die Erfolgslage verstärkt vor der Vermögens- und Finanzlage zu analysieren:

- * beschaffen mehr Informationen,
- * nutzen mehr aggregierte Daten und
- * bilden mehr Kennzahlen.

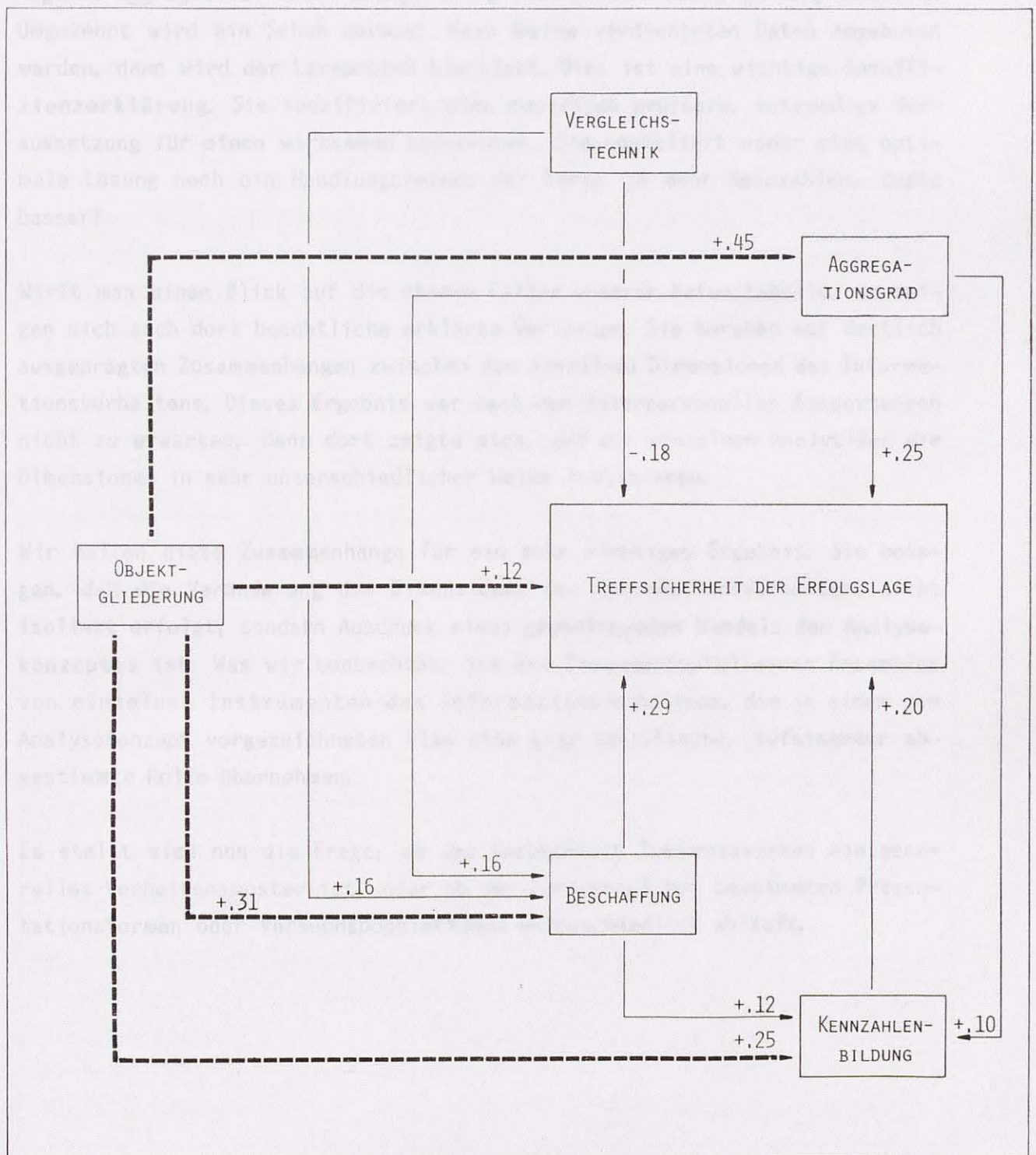
Zur Bestimmung der Stärke dieser drei Teilwirkungen werden die direkten Einflüsse der Objektgliederung auf die jeweiligen Verhaltensdimensionen mit deren totalem Effekt multipliziert. Es ergeben sich folgende Teilwirkungen:

1. Verstärkte Informationsbeschaffung:	$0,31 * 0,31 = 0,10$
2. Intensivere Nutzung aggregierter Daten:	$0,45 * 0,32 = 0,14$
3. Verstärkte Kennzahlenbildung:	$0,25 * 0,20 = 0,05$

Summe der wichtigsten indirekten Effekte	0,29
--	------

Der Pfad über die **intensivere Nutzung aggregierter Daten** ist offensichtlich am bedeutsamsten für den Lernfortschritt. Dies bedeutet, daß der Entwicklung problemgerechter Informationsangebote eine Schlüsselrolle für die Steigerung der Effizienz bilanzanalytischer Informationsprozesse zukommt. Berücksichtigt man ferner, daß man einen Teil der Kennzahlenbildung an den Computer delegieren kann, insbesondere bei flexiblen Lösungen, die eine Anpassung an persönlichkeitsabhängige Informationsbedürfnisse ermöglichen, dann kommt der maschinellen Unterstützung des bilanzanalytischen Informationsprozesses eine außerordentlich wichtige Funktion zu.

Abbildung 4.27.: Intertemporale Analyse – Einflüsse der Objektgliederung auf die Treffsicherheit der Erfolgslage (Basis: Je 412 ipsative Meßwerte)



Vor einer zu mechanistischen Interpretation der Zusammenhänge muß jedoch gewarnt werden. Man sollte den Befund nicht so lesen, daß ein größeres Angebot aggregierter Daten zwangsläufig eine Effizienzsteigerung bewirkt. Umgekehrt wird ein Schuh daraus: Wenn **keine** verdichteten Daten angeboten werden, dann wird der Lernprozeß blockiert. Dies ist eine wichtige **Insuffizienzenerklärung**. Sie spezifiziert eine empirisch prüfbare, notwendige Voraussetzung für einen wirksamen Lernprozeß. Sie postuliert weder eine optimale Lösung noch ein Handlungsrezept der Form: je mehr Kennzahlen, desto besser!

Wirft man einen Blick auf die oberen Felder unserer Befundtabelle, so zeigen sich auch dort beachtliche erklärte Varianzen. Sie beruhen auf deutlich ausgeprägten Zusammenhängen zwischen den einzelnen Dimensionen des Informationsverhaltens. Dieses Ergebnis war nach den interpersonellen Auswertungen nicht zu erwarten, denn dort zeigte sich, daß die einzelnen Analytiker die Dimensionen in sehr unterschiedlicher Weise kombinieren.

Wir halten diese Zusammenhänge für ein sehr wichtiges Ergebnis. Sie besagen, daß die Veränderung der Dimensionen des Informationsverhaltens nicht isoliert erfolgt, sondern Ausdruck eines **grundlegenden Wandels des Analysekonzeptes** ist. Was wir beobachten, ist das **Zusammenspiel** eines Ensembles von einzelnen Instrumenten des Informationsverhaltens, die in einem vom Analysekonzept vorgezeichneten Plan eine ganz spezifische, aufeinander abgestimmte Rolle übernehmen.

Es stellt sich nun die Frage, ob das beobachtete Zusammenwirken ein generelles Verhaltensmuster ist, oder ob der Lernprozeß bei bestimmten Präsentationsformen oder Versuchspopulationen unterschiedlich abläuft.

Alle indirekten Effekte	0,28	0,29	0,25	0,30
Direkter Effekt	0,12	0,22	0,17	0,02
Totaler Effekt	0,40	0,45	0,42	0,40

4.3.2.5.2. Differenzierung nach Präsentationsformen: Bestätigung der Verhaltensmuster

Tabelle 4.21. belegt die **Stabilität der Einflüsse**: Der positive totale Effekt der Objektgliederung tritt bei allen drei Präsentationsformen auf, und zwar in fast identisch gleicher Stärke! Auch die Einflußstruktur ist sehr ähnlich: die Einflußpfade über die intensivere Nutzung aggregierter Daten und die verstärkte Kennzahlenbildung sind bei allen drei Präsentationsformen praktisch gleich ausgeprägt. Nur bei der Informationsbeschaffung zeigt sich eine deutliche Abweichung. Sie beruht auf den unterschiedlichen direkten Effekten der Informationsbeschaffung bei den einzelnen Präsentationsformen.

Tabelle 4.21.: Intertemporale Analyse – Einflüsse der Objektgliederung auf die Treffsicherheit der Erfolgslage (Basis: 412 ipsative Meßwerte)

Effekte	Alle Prozesse	Tabelle	Baum nur bei Analyse	Baum bei Analyse u. Vortrag
Wichtigste indirekte Effekte				
Informationsbeschaffung	0,12	0,08	0,04	0,18
Aggregationsgrad	0,14	0,12	0,16	0,14
Kennzahlenbildung	0,05	0,04	0,05	0,05
Alle indirekten Effekte	0,28	0,24	0,25	0,38
Direkter Effekt	0,12	0,22	0,17	0,02
Totaler Effekt	0,40	0,45	0,42	0,40

Tabelle 4.22. zeigt diese Unterschiede. Der direkte Effekt beträgt bei der tabellarischen Informationspräsentation 0,11, bei der nur während der Fallbearbeitungen präsentierten Baumstruktur 0,36 und bei der Baumstruktur, die während der Fallbearbeitung und während des Vortrags verwendet wird, 0,45. Die Baumstruktur lenkt die Informationssuche auf fallspezifisch relevante Informationen. Daß der Effekt bei der Serie Kiel 84 besonders stark ausgeprägt ist, liegt daran, daß bei ihr die Vorzüge der Baumstruktur in "Wort und Bild" erläutert werden.

Die weiteren Befunde zeigen bei allen drei Präsentationsformen einen **negativen Einfluß der Verrichtungsgliederung**. Dies bestätigt unsere Zweifel gegenüber einer **starren** Phasengliederung. Anzumerken ist, daß sich unser Befund nur auf die Informationen der Erfolgslage, d. h. bereits auf eine Teilmenge aller Informationen bezieht. Auch **innerhalb** dieses Analysefelds ist es offenkundig besser, kleinere Informationspakete zu schnüren und nach inhaltlich abgegrenzten **Analysemodulen** vorzugehen, um überschaubare Teilprobleme zu generieren, die auf markante Schlüsselergebnisse reduziert werden können. Die prozessualen Studien anderer Forscher belegen, daß eine derartige Vorgehensweise von Experten angewandt wird, weil sie ein kapazitätsentlastendes "chunking" ermöglicht. Aus diesem Blickwinkel ist es nicht so verwunderlich, daß sich die Analytiker mit zunehmender Erfahrung immer stärker von einem postulierten phasentheoremkonformen Verhalten lösen, statt sich diesem anzunähern.

Auch bezüglich der übrigen totalen Effekte gibt es eine bemerkenswerte Parallelität der Befunde: Bei allen drei Präsentationsformen bewirken die intensivere Nutzung aggregierter Daten und die verstärkte Kennzahlenbildung eine Effizienzsteigerung. Die Übertragungsaktivität zeigt bei keiner der drei Präsentationsformen einen nennenswerten Effekt. Bei der Vergleichstechnik ergeben sich in allen drei Fällen negative Einflüsse, nur das Ausmaß ist bei der Serie Kiel 84 besonders stark. Wie die Detailanalyse zeigt, liegt dies daran, daß die Teilnehmer dieser Serie in besonders starkem Maße Betriebsvergleiche reduziert haben.

Tabelle 4.22.: Intertemporale Analyse – Pfadanalyse zur Treffsicherheit der Erfolgslage differenziert nach Präsentationsformen

Präsentationsform Tabelle (Basis: Je 132 ipsative Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korre-lation
Verrichtungs-Gliederung	-0.04	-0.16	-0.12	-0.13
Objekt-Gliederung	0.22	0.45	0.24	0.44
Vergleichstechnik	-0.10	-0.06	0.04	-0.07
Aggregationsgrad	0.29	0.35	0.06	0.47
Beschaffung	0.11	0.16	0.05	0.39
Übertragung	-0.03	-0.01	0.02	0.18
Kennzahlenbildung	0.24	0.24	----	0.44

Präsentationsform Baum, nur bei Analyse (Basis: Je 136 ipsative Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korre-lation
Verrichtungs-Gliederung	-0.03	-0.10	-0.07	-0.09
Objekt-Gliederung	0.17	0.42	0.25	0.42
Vergleichstechnik	-0.27	-0.09	0.18	-0.11
Aggregationsgrad	0.30	0.38	0.07	0.44
Beschaffung	0.36	0.37	0.01	0.40
Übertragung	0.00	0.02	0.02	0.06
Kennzahlenbildung	0.13	0.13	----	0.36

Präsentationsform Baum, bei Analyse und Vortrag
(Basis: Je 144 ipsative Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korre-lation
Verrichtungs-Gliederung	-0.03	-0.24	-0.21	-0.29
Objekt-Gliederung	0.02	0.40	0.38	0.43
Vergleichstechnik	-0.25	-0.24	0.01	-0.30
Aggregationsgrad	0.15	0.23	0.08	0.44
Beschaffung	0.45	0.47	0.02	0.58
Übertragung	-0.05	-0.02	0.04	0.08
Kennzahlenbildung	0.19	0.19	----	0.44

Insgesamt gesehen zeigt der Befund sehr stabile, reproduzierbare Effekte. Dies spricht für unsere Dateninterpretation. Zu prüfen bleibt, ob die Effekte auch bei einer Differenzierung nach Populationen gelten.

4.3.2.5.3. Differenzierung nach Populationen: Erneute Bestätigung der Verhaltensmuster

Konzentriert man sich auf die in Tabelle 4.23. ausgewiesenen totalen Effekte, so ist folgendes festzustellen:

1. Bei vier der sieben Dimensionen ergeben sich übereinstimmend bei allen drei Populationen deutliche positive Effekte (Objektgliederung, Aggregationsgrad, Beschaffung, Kennzahlenbildung).
2. Veränderungen der Übertragungsaktivität beeinflussen bei keiner der drei Serien die Entwicklung der Effizienz.
3. Eine Vereinfachung der Vergleichstechnik wirkt sich bei allen drei Serien förderlich auf die Entwicklung der Effizienzmeßwerte aus. Die Effekte sind bei den Kieler Teilnehmern stärker als bei den Gießener Analytikern.
4. Abweichungen von einer phasentheoremkonformen Abfolge der Verrichtungen sind bei der Serie Gießen 82 unbeachtlich für die Effizienz der Analyse. Bei den Serien Kiel 82 und Kiel 84 ist die Effizienz um so höher, je stärker sich die Abfolge von der postulierten Norm entfernt.

Insgesamt gesehen bedeutet dieses Ergebnis eine Bestätigung des Strukturmodells. Unterschiede im Ausmaß einzelner Koeffizienten und die Abweichungen bei der Verrichtungsgliederung könnten zum Gegenstand von weiterführenden Detailanalysen gemacht werden. Davon wollen wir absehen, weil das grundsätzliche Ziel, die Stabilität und Reproduzierbarkeit des Verhaltensmusters zu überprüfen, erfüllt ist.

Tabelle 4.23.: Intertemporale Analyse – Pfadanalyse zur Treffsicherheit der Erfolgslage differenziert nach experimentellen Serien

Gießen 82 (Basis: Je 136 ipsative Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korrelation
Verrichtungs-Gliederung	0,07	-0,02	-0,09	0,06
Objekt-Gliederung	0,14	0,47	0,34	0,47
Vergleichstechnik	-0,22	-0,08	0,14	-0,01
Aggregationsgrad	0,38	0,45	0,07	0,53
Beschaffung	0,27	0,29	0,02	0,49
Übertragung	0,00	0,02	0,02	0,20
Kennzahlenbildung	0,15	0,15	----	0,37

Kiel 82 (Basis: Je 132 ipsative Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korrelation
Verrichtungs-Gliederung	-0,22	-0,34	-0,12	-0,40
Objekt-Gliederung	0,17	0,31	0,14	0,37
Vergleichstechnik	-0,15	-0,15	0,00	-0,19
Aggregationsgrad	0,20	0,24	0,04	0,34
Beschaffung	0,18	0,18	0,00	0,29
Übertragung	-0,09	-0,08	0,01	-0,02
Kennzahlenbildung	0,18	0,18	----	0,43

Kiel 84 (Basis: Je 144 ipsative Meßwerte)

Unabhängige Variablen	Direkter Effekt	Totaler Effekt	Indir. Effekt	Korrelation
Verrichtungs-Gliederung	-0,03	-0,24	-0,21	-0,29
Objekt-Gliederung	0,02	0,40	0,38	0,43
Vergleichstechnik	-0,25	-0,24	0,01	-0,30
Aggregationsgrad	0,15	0,23	0,08	0,44
Beschaffung	0,45	0,47	0,02	0,58
Übertragung	-0,05	-0,02	0,04	0,08
Kennzahlenbildung	0,19	0,19	----	0,44

4.3.2.6. Zusammenfassende Würdigung der intertemporalen Analyse

1. Unsere **Hypothesen**, die einen **umgekehrt u-förmigen Zusammenhang** von **Informationsverhalten** und **Effizienz** postulieren, sind auch in der intertemporalen Analyse zweifelsfrei **widerlegt** worden.
2. Bei der **Erfolgslage** zeigt sich eine deutliche **Verbesserung der Effizienz der Analyse**, bei der Vermögens- und Finanzlage ist eine Verschlechterung festzustellen. Der positive Effekt bei der Erfolgslage ist jedoch wesentlich stärker, so daß sich **insgesamt eine Leistungssteigerung** bezüglich der Treffsicherheit der Beurteilung der Gesamtlage ergibt.

Die Analyse konzentriert sich daher auf das **Lernverhalten bei der Erfolgslage**. Für sie gelten die nachfolgenden Aussagen.

3. Die intertemporale Analyse **bestätigt** den **positiven Einfluß der Informationsverdichtung**: die intensivere Nutzung aggregierter Daten und die vermehrte Bildung von Kennzahlen stehen in einer signifikant positiven Beziehung zur beobachteten Effizienzsteigerung. Hieraus kann die praxeologische Konsequenz gezogen werden, daß gute Analytiker durch maschinelle Informationssysteme wirksam unterstützt werden können, denn ohne verdichtete Daten analysieren sie schlechter, werden ihre Lernprozesse gehemmt.
4. Die Verbesserung der Effizienz während der Blockveranstaltung ist mehr als eine Summe von Teileinflüssen. Unsere Pfadanalyse zeigt vielmehr einen **grundlegenden Wandel des Analysekonzeptes** an, bei dem die einzelnen Instrumente des Informationsverhaltens in einem aufeinander abgestimmten Mix verändert werden. Ausgangspunkt dieses Wandels ist die Veränderung der **Ablauforganisation** des Informationsprozesses. Die Analytiker lösen sich von dem Vorbild der bilanzanalytischen Literatur und beginnen mit der Erfolgslage statt mit der Vermögens- und Finanzlage. Die systematische Erfolgsquellenzerlegung hat eine effizienzsteigernde intensivere Nutzung aggregierter Daten, eine verstärkte Informationsbeschaffung und eine vermehrte Kennzahlenbildung zur Folge. Alle drei Teilwirkungen zusammen führen zu einem deutlichen indirekten positiven

Effekt der Ablauforganisation.

5. **Der positive Effekt des Analysekonzeptes erweist sich als bemerkenswert stabil und reproduzierbar.** Wenn man die Befunde nach Präsentationsformen und Populationen differenziert, bleibt die Einflußstruktur erhalten. Das beobachtete Verhaltensmuster stellt demnach keinen einmaligen Effekt dar, sondern eine generalisierbare Folge des verwendeten bilanzanalytischen Trainingsprogramms.
6. Die bei der interpersonellen Analyse festgestellten negativen Effekte der **Übertragungsaktivität** gelten **nicht** in dem Maße im intertemporalen Vergleich, obwohl die Menge an übertragenen Informationen durchaus verändert wird. Dies ist ein überraschendes, bisher noch nicht erklärtes Phänomen. Wir hatten erwartet, daß die Übertragungsaktivität wegen ihrer hohen Kapazitätsbelastung effizienzsenkend wirkt.
7. Die **Komplexität der Vergleichstechnik** übt auch in der intertemporalen Analyse einen negativen Einfluß auf die Effizienz aus. Die Entwicklung der Meßwerte dieser Verhaltensdimension legt jedoch eine andere Interpretation nahe: Die Effizienz **verbessert** sich, weil die Analytiker dazu übergehen, eine einfachere, ihrer beschränkten Kapazität besser angepaßte Vergleichstechnik anzuwenden. Dieser positive Effekt ist darauf zurückzuführen, daß nur noch die Informationen beschafft und übertragen werden, die man tatsächlich (sachgerecht) nutzen kann.